

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE



BUILDING THE 21ST CENTURY SCIENTIST

Why science
education
is more
important
than ever

PAGE 271

DEVELOPMENTAL BIOLOGY

STEM CELLS TAKE THE TUBE

Cell-cell communication
via natural nanotubes

PAGES 292 & 329

GEOCHEMISTRY

MAKING CONTINENTS

The formation of Earth's
continental crust

PAGES 293 & 301

ASTROCHEMISTRY

BUCKYBALLS IN SPACE

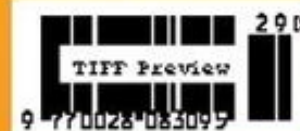
Fullerene C₆₀ detected
in the Milky Way

PAGES 296 & 322

NATURE.COM/NATURE

16 July 2015 £10

Vol. 523, No. 7560

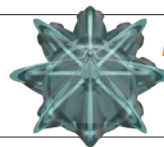


THIS WEEK

EDITORIALS

EDUCATION The changing needs of science pupils and teachers **p.256**

WORLD VIEW Indonesian research shows danger of national pride **p.257**



ALL SET Ancient comb jellies had rigid spokes and skeleton **p.258**

Unholy alliance

An independent report on the American Psychological Association reveals the extent to which some psychologists colluded with US military and intelligence agencies to allow torture of prisoners.

In 1917, when the field of psychology was young and struggling to gain acceptance in science, the American Psychological Association (APA) needed a friend. Like many at the time, it decided to assist the war effort by working with the US military. The collaboration was largely benign: efforts to assess which recruits were fit to be soldiers led to the first formal study of variation in human intelligence. Later, psychologists studied the effects of war on soldiers returning home, fuelling the case for making the First World War “the war to end all wars”.

That was not to be, but psychology, and the APA in particular, continued its close bond with military and intelligence agencies. The relationship is not inherently problematic: indeed, the US Department of Defense (DOD) spends tens of millions of dollars each year on research into post-traumatic stress disorder and other psychological and psychiatric complications of war. The DOD, which employs around 700 psychologists, was a key ally in psychologists obtaining the authority to write prescriptions in the 1980s. And the APA has at times taken a stand against DOD policies: as early as 1991, the organization protested against the Pentagon’s policy of stopping openly gay people from serving in the military by banning DOD advertisements in APA publications.

Nevertheless, the tone of the alliance between US agencies and psychologists has darkened over the past century. Most famous is the Central Intelligence Agency’s (CIA) mind-control programme MKUltra during the cold-war era, in which psychologists helped the CIA to develop and test interrogation techniques involving the use of hallucinogenic drugs and hypnosis.

Given this history, it should be no surprise that the APA has continued to cultivate a close relationship with the agencies. Last week, a long-awaited external report confirmed suspicions of the APA’s involvement in the torture of detainees following the terrorist attacks of 11 September 2001 and the ensuing ‘war on terror’ (see go.nature.com/4vpdob). Starting in 2005, the report found, APA officials worked with the DOD to keep the organization’s ethics guidelines loose enough to justify the participation of psychologists in the DOD’s ‘enhanced interrogation’ programme. As a result, the DOD and CIA could easily brush aside the ethical concerns of their own psychologists: the APA had given the programme its imprimatur.

The story is rife with conflicts of interest: according to the report, six of the nine voting psychologists on the APA task force that wrote the guidelines had consulting relationships with the DOD or CIA, and one former APA president owned a financial stake in the consulting company that oversaw the CIA interrogation programme. As criticism surfaced, the APA defended itself by formally condemning torture while doing nothing to stop its members from participating. Meanwhile, Guantanamo Bay’s chief military psychologist told an APA meeting: “If we removed psychologists from these facilities, people are going to die.” It is an assertion that does more to reveal the disgraceful state of the programme than to offer a moral defence.

Not only did APA psychologists deem the torture programme

ethical, but they also gave it a patina of legitimacy by trying to cast it as research. The “studies” — which violate every consent rule for human subjects, including the CIA’s own — involved questions about the acceptable limits of human suffering and how well various techniques could yield useful information from a prisoner. There is no evidence that the United States gained any useful information in this way.

The scientific basis for the interrogation programme was questionable from the start. The theory of ‘learned helplessness’ was developed to test psychiatric drugs by measuring how long mice will try to swim in a bucket of water — depressed animals will give up sooner and allow themselves to be rescued. The psychologists who developed the CIA’s interrogation techniques reversed this idea, theorizing that simulating the experience of drowning, or waterboarding, could induce despair in human detainees until they gave up their story.

The APA has apologized for its failings and has indicated that it will revise its policies to prohibit psychologists from participating in military interrogations. It has also parted company with its ethics director, who the report named as leading the collusion with the military. More heads are likely to roll, and some psychologists could even face prosecution.

The American Psychiatric Association and the American Medical Association forbade their members in 2006 from participating in the interrogation programmes. This is in keeping with the Geneva Convention, an international agreement signed in 1929 and revised nearly 70 years ago to do away with torture and abusive experimentation on prisoners of war. The APA deserves all the criticism it receives and more, for its willingness to forswear global consensus in the interest of making a deal with the devil. ■

“There is no evidence that the United States gained any useful information.”

Austerity bites

If the UK government is serious about science, now is the time to prove it.

Guess the year. With a major international ‘make or break’ climate summit scheduled to take place in a European capital, campaigners protest against plans to build a third runway at London’s Heathrow airport. Greece faces a debt crisis, prompting political upheaval across the continent and fears for the future of the euro currency. Serena Williams and Roger Federer play in their respective finals of the Wimbledon tennis championships in London.

Plus ça change. That was — in fact — 2009, but it describes 2015

equally well. And the two years have something else in common: political investment in science sits at a crossroads.

On 9 July, a group of scientists set up to advise the United Nations secretary-general Ban Ki-moon startled many researchers with a bold assertion: nations should invest up to 3.5% of their gross domestic product (GDP) in science.

Cue snorts of derision. Although a tiny group of nations invests around this much — Sweden and Israel among them — most fall well below this threshold. According to the latest figures from the Organisation for Economic Co-operation and Development, the United States invests 2.7%, and China 2%. The European Union average comes in at just under 2%.

Even the UN science advisory board admits that a target of 1% is perceived as high by many governments. It does, however, say that 3.5% of GDP is necessary to put the world on a sustainable development course. If this target seems rather arbitrary, it is because it probably is. But this crude measure of support for science can still be a useful metric.

Take the case of the United Kingdom. Combined private and public spending on UK science is around 1.6% of GDP. Earlier this year, the heads of various learned societies called for politicians to increase this figure to 3%, but the plea raised little more than eyebrows.

An ambition to boost government spending on science might have received a more welcome response in 2009 — but since then austerity has dominated in the United Kingdom. The Conservative–Liberal Democrat coalition government that came to power in 2010 did fulfil its promise to protect the core UK science budget from cuts, but inflation has whittled away the amount that is available for research.

Following last week's UK budget statement, there are signs that austerity measures are being relaxed — for some at least. In the first fiscal plan produced by a majority Conservative government for nearly two decades, Chancellor of the Exchequer George Osborne announced some cuts — to welfare benefits and national broadcaster the BBC, for example — but he also unveiled significant

belt-loosening measures, including tax cuts for the middle classes.

Exactly what this means for science is not yet clear. The Conservatives say that they will cut about £17 billion (US\$26 billion) from government departments. Some of these axe blows may fall on research spending.

But the party has been vocal in its support for some scientific projects. They have championed the (nebulous) term 'innovation' as key to improve the Britain's woeful workplace productivity. And cash has flowed, up to a point, to huge projects such as the Francis Crick Institute for biomedical research in London and the National Graphene Institute in Manchester.

Still, of Britain's 1.6% of GDP spent on science, the public spend makes up just 0.44%. Compare that with Germany, where the government contributes 0.85% of GDP out of an overall spend on science of 2.9% of GDP. And the US government spends 0.76% of GDP out of an overall investment in science of 2.7% of GDP.

If Osborne is serious about science, now is the time to prove it. At a parliamentary gathering last month, at which politicians rubbed shoulders with researchers, the subject of science funding was on the lips of many. A reference to the percentage of GDP spent on science has become de rigueur in such conversations, often with an addendum that the United Kingdom 'punches above its weight' in achieving what it does with its limited means. This attitude has almost become part of the political identity of UK science: 'we do so well with so little — why not give us more money and let us show you what we can really do'.

It has a point — if there is money to cut taxes, there should be money to support the work that can drive economies.

There are, of course, many claims on public financing, and scientists must be prepared to fight for their share alongside hospital administrators, road builders and arts funders. But if the UK government wishes to continue to wear the mantle of a science supporter, pushing towards 3.5% would be a step in the right direction. ■

An education

A special issue looks at how science is taught — and why a change in methods is essential.

One of the subjects that people love to argue about, following closely behind the 'correct' way to raise children, is the best way to teach them. For many, personal experience and centuries of tradition make the answer self-evident: teachers and textbooks should lay out the content to be learned, students should study and drill until they have mastered that content, and tests should be given at strategic intervals to discover how well the students have done.

And yet, decades of research into the science of learning has shown that none of these techniques is particularly effective. In university-level science courses, for example, students can indeed get good marks by passively listening to their professor's lectures and then cramming for the exams. But the resulting knowledge tends to fade very quickly, and may do nothing to displace misconceptions that students brought with them.

Consider the common (and wrong) idea that Earth is cold in the winter because it is further from the Sun. The standard, lecture-based approach amounts to hoping that this idea can be displaced simply

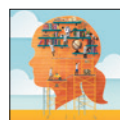
by getting students to memorize the correct answer, which is that seasons result from the tilt of Earth's axis

of rotation. Yet hundreds of empirical studies have shown that students will understand and retain such facts much better when they actively grapple with challenges to their ideas — say, by asking them to explain why the northern and southern hemispheres experience opposing seasons at the same time. Even if they initially come up with a wrong answer, to get there they will have had to think through what factors are important. So when they finally do hear the correct explanation, they have already built a mental scaffold that will give the answer meaning.

In this issue, prepared in collaboration with *Scientific American*, *Nature* is taking a close look at the many ways in which educators around the world are trying to implement such 'active learning' methods (see page 271). The potential pay-off is large — whether it is measured by the increased number of promising students who finish their degrees in science, technology, engineering and mathematics (STEM) disciplines instead of being driven out by the sheer boredom of rote memorization, or by the non-STEM students who get first-hand experience in enquiry, experimentation and reasoning on the basis of evidence.

Implementing such changes will not be easy — and many academics may question whether they are even necessary. Lecture-based education has been successful for hundreds of years, after all, and — almost by definition — today's university instructors are the people who thrived on it.

But change is essential. The standard system also threw away far too many students who did not thrive. In an era when more of us now work with our heads, rather than our hands, the world can no longer afford to support poor learning systems that allow too few people to achieve their goals. ■



THE 21ST CENTURY SCIENTIST
A *Nature* and *Scientific American*
special issue nature.com/stem

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqv



Don't distort policy in the name of national pride

Dyna Rochmyaningsih offers a lesson from Indonesia on what can go wrong when governments use research to make a country look good.

Countries have long tried to capitalize on scientific achievements by celebrating them as a source of national pride. From the state sponsorship of polar expeditions a century ago to the state-sponsored programme earlier this year from the Iranian government that its nuclear programme was developed to gain prestige, politicians recognize the powerful and popular pull of success in research. In a 2002 survey, for example, more people in the United States said that they were “very proud” of the nation’s achievements in science and technology than in any other field, including sport, economics, and art and literature. And in 2013, the BBC devoted a series of television documentaries to the “Wonder of British science”.

For a developing nation such as Indonesia, advancing its science and technology is therefore a convenient way for the government to encourage its people to feel better about their country and its place in the world. But what happens when nationalism drives, rather than merely celebrates, science? The results from Indonesia suggest that priorities can become skewed and rigorous scientific assessment ignored. This is a valuable lesson for all countries.

The development of science and technology in Indonesia has accelerated in recent years, with help from US scientific diplomacy in the Islamic world. In May, the Indonesian Academy of Sciences, with support from the United States, launched a fund offering grants to young researchers. Collaborating with Australia, the academy has also set a series of challenges for Indonesian scientists to address, including disaster mitigation and strengthening community resilience.

Such international support for Indonesian science seems stable — the partnership with Australia survived the political row between the two countries in April over Indonesia’s execution of Australian drug smugglers, for example. Yet although it is true that science has no boundaries, strong and original research by a country such as Indonesia is a crucial platform on which to build its economic development. Indonesia’s scientists seem capable: most of the country’s patent applications, for example, come from scientists in local universities. But in what direction does the Indonesian government want them to head?

Two high-profile projects suggest that the officials who steer our national science policy are favouring questionable research that is likely to bring short-term headlines and ‘national pride’, rather than solid science that will result in true, long-term societal benefit.

The first of these projects is based on the claim that ancient Indonesia could have been home to Earth’s oldest civilization. Geological patterns discovered during excavation of the Gunung Padang volcanic site in West Java have been used to push the idea that an ancient pyramid is buried there that would

predate the historic ruins of ancient Egypt by thousands of years. Many Indonesian people love the project and its extraordinary claims, and the work has received backing from the government. The Indonesian president visited the site last year to proclaim it the cradle of civilization, and military forces have been deployed to assist work there.

Ironically, Danny Hilman Natawidjaja — the lead scientist on the project and a geologist at the Indonesian Institute of Sciences — was originally studying evidence of past natural disasters in the region, as part of research to understand future hazards, when he became distracted by the pyramid idea. Nationalism has distorted priorities. Studying disasters is more important than searching for a buried pyramid.

The second project concerns a controversial electrical cancer therapy, developed and promoted by Warsito Taruno at CTECH Laboratories

in Tangerang. Devices based on static electricity, he claims, can locate and then target cancer cells to stop them dividing. Indonesian people with cancer are flocking to his clinic for the treatment.

When the Indonesian Radiation Oncology Society and the Indonesian Society of Surgical Oncology sent a letter to the government in 2013 asking it to close Warsito’s clinic — on the grounds that the device had not been clinically trialled in Indonesian hospitals — the government again played the national-pride card. The Ministry of Health said that Warsito’s research is innovation from a “child of the nation” and should be supported. And last year, the Ministry of Legal and Human Rights honoured him as our country’s best inventor.

If the government is serious about building a future for Indonesia on the basis of devices such as Warsito’s, it should carry out further human testing and double-blind clinical trials. In its response to the oncology associations’ letter, it said that it was willing to do this. More broadly, no government should base scientific development on what it thinks makes its country look good. Our government does support solid research that could help Indonesia to flourish using its national resources, such as projects to investigate the medicinal properties of native plants. But because these projects receive little public attention, officials do not view such efforts as a priority.

Fixing the problem will not be easy. Perhaps a science adviser to the president would help. Or science policy might be centralized in a single ministry. The Indonesian Academy of Sciences could play a more active part, by speaking out to promote high-quality science. And most of all, the government must realize that science is more than a source of national pride. Its ultimate goal is not to put a smile on people’s faces. It is to help them to live better lives. ■

Dyna Rochmyaningsih is a freelance science journalist in Jakarta.
e-mail: drochmya87@gmail.com

STUDYING
DISASTERS
IS MORE
IMPORTANT
THAN SEARCHING
FOR A BURIED
PYRAMID.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/qtamma

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

FLUID DYNAMICS

Many microbes make a superfluid

Bacteria swimming in a fluid can reduce its viscosity to zero.

Unlike most liquids, superfluids have zero viscosity and so can behave in an unusual way, sometimes even defying gravity. Harold Auradou at the University of Paris-Sud in France and his colleagues found that they could alter the viscosity of a normal liquid solution by varying the concentration of *Escherichia coli* swimming in it. As the researchers spun the solution around in a rheometer (which measures viscosity), the motion of the bacteria's propulsive tails compensated for the forces between the molecules in the liquid, decreasing its viscosity. Once there were around 6 billion strong swimmers in suspension, the viscosity fell to zero.

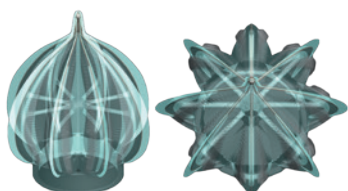
Phys. Rev. Lett. 115, 028301 (2015)

EVOLUTIONARY BIOLOGY

Ancient jellies had skeletons

Fossils of comb-jelly ancestors show signs of rigid, skeletal spokes and plates, unlike their living, soft-bodied descendants.

All 150 known species of modern comb jellies, known as ctenophores, lack skeletons and use eight rows (or combs) of hair-like structures to swim. Qiang Ou of the China University of Geosciences in Beijing and his colleagues



identified three new species of comb jelly from roughly 520-million-year-old Cambrian deposits in south China, and reanalysed these and other comb-jelly fossils. The fossils had combs and a similar basic body plan to living ctenophores, but also had radiating spokes and rigid plates (reconstructed models **pictured**), which probably provided support or served as armour.

The intense interactions of ancient comb jellies with their predators and prey

may have led to this burst of diverse body types in the Cambrian period, the authors suggest.

Sci. Adv. 1, e1500092 (2015)

CLIMATE CHANGE

Changing winds mean longer flights

Shifting wind patterns caused by climate change could lengthen some airline flights, further raising carbon dioxide emissions.

Kristopher Karnauskas at

Woods Hole Oceanographic Institution in Massachusetts and his colleagues analysed the duration of 250,000 flights over the North Pacific and found that flight times varied with natural changes in wind speed at cruising altitude. Using climate models, the team predicts that climate change driven by greenhouse gases could cause part of the Pacific jet stream to shift into the corridor between Hawaii and the west coast of the continental United States, increasing average flight



CLIMATE SCIENCE

Black Sea warming caused extreme rain

Intense rainfall that caused a devastating flash flood in a Russian town in 2012 (pictured) has been linked to the increasing surface temperature of the Black Sea.

The flood in the town of Krymsk killed more than 170 people after an unprecedented amount of rain — twice the previous record — fell in a single night. Edmund Meredith at the GEOMAR Helmholtz Centre for Ocean Research in Kiel, Germany, and his co-workers used an atmospheric model to study the

sensitivity of this event to Black Sea warming. They found that simulations using current Black Sea surface temperatures produced a more than 300% increase in rainfall compared with models in which sea temperatures were similar to those of the early 1980s.

The warmer waters destabilize the atmosphere, making it more likely that intense thunderstorms will form.

Nature Geosci. <http://dx.doi.org/10.1038/ngeo2483> (2015)

MIKHAIL MORDASOV/AFP/GETTY

QIANG OU ET AL.

times on this route.

The authors warn that such changes would increase fuel consumption and CO₂ emissions on this route and possibly for other flights globally, which would further drive climate change.

Nature Clim.

Change <http://dx.doi.org/10.1038/nclimate2715> (2015)

CLIMATE-CHANGE BIOLOGY

Bees squeezed by warming

Bumblebee ranges are shrinking as the world warms, suggesting that climate change is yet another factor threatening the pollinators.

A team led by Jeremy Kerr at the University of Ottawa, Canada, collated more than 420,000 observations of bumblebee species from North America and Europe between 1975 and 2010. They found that as temperatures rose, the southern limits of many bumblebee species' ranges moved north, by as much as 300 kilometres in some cases. But the northern edges of the bees' ranges stayed in place, leading to an overall contraction of the insects' habitats.

Neither pesticide usage nor changes in land use correlated with the observed shifts in the bumblebee ranges.

Science 349, 177–180 (2015)

BEHAVIOURAL ECOLOGY

Why bats like to perch in a pitcher

Pitcher plants in Borneo have evolved specialized structures that attract bats, which roost inside the plants and fertilize them with their faeces.

Michael Schöner of the University of Greifswald in Germany and his colleagues bounced sonar waves off the rear inner wall of the pitcher plant *Nepenthes hemsleyana*



and found that it reflected the sound much more clearly and strongly than

did other species of pitcher plant. When they modified or removed the sound reflector, wild Hardwicke's woolly bats (*Kerivoula hardwickii*; pictured) took much longer to find the plant.

The authors suggest that such acoustically mediated symbiotic relationships may be more common than was thought.

Curr. Biol. <http://doi.org/56c> (2015)

GENE THERAPY

Deafness can be reversed in mice

Researchers have partially restored the hearing of deaf mice by delivering functioning genes into their ear cells.

Jeffrey Holt at Boston Children's Hospital in Massachusetts and his colleagues tested deaf mice that have mutations in the *Tmc1* gene, which causes 4–8% of genetic deafness in some human populations. They injected the mice with a virus carrying the gene and found that hair cells in the inner ear — which normally convert sound vibrations into electrical signals — took up and expressed the gene. The animals showed startle reflexes and shown responses to sounds.

This approach could one day complement other hearing-loss therapies, such as cochlear implants and hearing aids, the authors say.

Sci. Transl. Med. 7, 295ra108 (2015)

PHYSIOLOGY

How bear bones stay strong

Bones weaken with inactivity in most mammals, but hibernating bears maintain theirs by suppressing

SOCIAL SELECTION

Popular topics
on social media

US postdocs hope for overtime pay

A proposed regulation by US President Barack Obama that would extend overtime pay to millions of workers triggered fierce discussion among academics — including some who think that it could result in heftier pay for postdocs (see go.nature.com/jgckvb). If approved, the regulation would enable salaried workers who earn less than about US\$50,400 per year to receive at least 1.5 times their usual rate for every extra hour worked beyond a 40-hour week. “This reform is a LABOR law. Give me A SINGLE compelling reason academics below 50K should not be treated like other workers in this case,” tweeted Francois Gould, an anatomy postdoc at Northeast Ohio Medical University in Rootstown, Ohio. But Belinda Huang, who is executive director of the

➔ **NATURE.COM**

For more on
popular papers:
go.nature.com/7viusy

National Postdoctoral Association, headquartered in Washington DC, says that it is unclear whether the proposal would apply to postdocs and other scientists.

bone turnover.

Seth Donahue at Colorado State University in Fort Collins and his colleagues analysed blood and bone samples from 13 black bears (*Ursus americanus*), which hibernate for up to 6 months every year. They found lower levels of key protein markers associated with bone formation and breakdown in hibernating bears than in active ones. The concentration of a hormone that reduces bone breakdown was 15 times higher during hibernation than during active periods.

Hibernating and active bears had the same level of calcium in their blood, suggesting that bears balance bone formation with breakdown during hibernation.

J. Exp. Biol. 218, 2067–2074 (2015)

ECOLOGY

Humans alter desert ecosystem

A community of small mammals in the western United States has changed more in the past century than in the past 13,000 years, thanks to human activities.

Rebecca Terry at Oregon State University in Corvallis and Rebecca Rowe at the University of New Hampshire in Durham analysed the skeletal remains of small mammals dating back to 12,800 years ago from a large cave in Utah. They found that energy flow — a metric combining the number, size and metabolic rates of animals — in this community was largely stable up until about the late 1800s, even during periods of rapid natural climate warming. But in the past 100 years, energy flow has dropped significantly, indicating a population surge in small animals that thrive in grassy habitats and a loss of animals in desert shrublands.

The authors suggest that this is a result of the replacement of shrublands with invasive grasses, owing to human activity, in this desert region.

Proc. Natl Acad. Sci. USA <http://dx.doi.org/10.1073/pnas.1424315112> (2015)

➔ **NATURE.COM**

For the latest research published by Nature visit:

www.nature.com/latestresearch

SEVEN DAYS

The news in brief

EVENTS

Torture collusion

The American Psychological Association (APA) released a long-awaited report on 10 July that concluded that its ethics guidelines on national security interrogations ultimately supported torture. The independent report, led by Chicago lawyer David Hoffman, found that APA officials colluded with the US Department of Defense (DOD), a major employer of psychologists, to craft permissive language for the guidelines that aligned with the DOD's own wishes. The APA released a statement apologizing for "organizational failures" and recommended updates to its ethical policies, such as prohibiting psychologists from participating in military or intelligence interrogations.

MERS lawsuit

Some of those affected by Middle East respiratory syndrome (MERS) in the large hospital-acquired outbreak in South Korea are suing the government and numerous hospitals. News of a lawsuit emerged on 9 July in the national newspaper *The Chosun Ilbo*.

NUMBER CRUNCH

£246 m

The estimated cost incurred by UK universities to participate in the Research Excellence Framework 2014 — a quality, impact and output assessment used by government funding agencies to distribute money. The figure (US\$384 million) was published in a review by Technopolis on 13 July.



KYE R. LEE/AP/PRESS ASSOCIATION IMAGES

Waco mammoths get monument status

The Waco Mammoth palaeontological site in Texas was one of three new areas designated as national monuments by US President Barack Obama on 10 July. The monument features the well-preserved remains of 24 Columbian mammoths (*Mammuthus columbi*) that lived about 65,000 years ago. They include the only

preserved nursery herd of mammoths in the United States. "This is one of the most incredible collections of mammoth fossils anywhere in the country," said Obama. The site will be protected under the Antiquities Act. The other new monuments are the Berryessa Snow Mountain in California and the Basin and Range in Nevada.

The plaintiffs, represented by the Citizens' Coalition for Economic Justice, argue that the government violated the country's constitution by failing to prevent a disaster and protect the public, and that the situation was made worse by its refusal to name affected hospitals. MERS infected 186 people in the outbreak, but no cases are known to have arisen since 4 July.

NASA soil probe

A radar instrument aboard a new NASA satellite to monitor soil moisture around the globe stopped transmitting on 7 July. The radar is one of two instruments aboard the Soil Moisture Active

Passive observatory, which launched in January. The satellite's mission is to measure moisture levels in the top 5 centimetres of soil to improve understanding of the links between Earth's water, energy and carbon cycles. NASA has convened a team to investigate.

Russian row

A private Russian science-funding organization, the Dynasty Foundation, is closing down after being branded a 'foreign agent' by Russia's Ministry of Justice. The foundation — which supports hundreds of young Russian researchers — said on its website that the decision

was taken at a board meeting on 5 July. Under a 2012 law, the label (which has connotations of spying), is reserved for non-governmental organizations that receive funding from abroad and are deemed to be involved in vaguely defined "political activities". See go.nature.com/jw7xbl for more.

FACILITIES

Telescope scrapped

The University of Hawaii at Hilo announced on 7 July that it will dismantle its 0.9-metre educational telescope on Mauna Kea, where native Hawaiians have been protesting against

the construction of another project, the Thirty-Meter Telescope (TMT). In May, the state's governor required the university to decommission one-quarter of the 13 telescopes on the mountain before the TMT is launched. The 0.9-metre instrument is the smallest of these and has not worked since it was installed in 2010.

RESEARCH

Rare gibbon family

A new family group of the world's rarest primates has been discovered. Last year, researchers warned that just 23–25 Hainan gibbons (*Nomascus hainanus*) remained (*Nature* **508**, 163; 2014). But a team led by the Zoological Society of London announced on 13 July that it had observed a previously unknown mating pair with a baby last month, adding not just three individuals to the count, but also a crucial extra social group.

Pluto sized up

On its way to a 14 July fly-by of Pluto, NASA's New Horizons probe discovered that the dwarf planet is 2,370 kilometres across, making it the largest-known object in the Solar System's icy Kuiper belt. That measurement is larger than many previous estimates for Pluto, making it bigger than



Eris, the 2,326-kilometre-wide dwarf planet that lies beyond it. New Horizons has also sent back a spectacular close-up of Pluto's surface (**pictured**), taken on the night of 13 July. See go.nature.com/dkztyn for more.

PEOPLE

Seabra steps down

Miguel Seabra has stepped down as president of research-advocacy group Science Europe with immediate effect. The Brussels-based group, which represents European research funders and institutions, announced on 10 July that Seabra had resigned less than a year into the job because of ill health. In April, he left his role as president of the Portuguese Foundation for Science and Technology. Civil engineer Elisabeth Monard, secretary-general of the Research Foundation Flanders, also in Brussels, will be acting president of Science Europe

until it elects a new president at its general assembly in November.

BUSINESS

Heart-drug approval

The US Food and Drug Administration has approved a new treatment for heart failure. The drug Entresto was approved on 7 July and is expected to be a blockbuster for its Swiss developer, Novartis. The drug combines two therapies: a previously approved treatment (sacubitril) for high blood pressure and heart failure, and a new drug (valsartan) that dilates blood vessels and promotes sodium excretion by inhibiting a protein called neprilysin. Entresto is the first approved neprilysin inhibitor.

POLICY

Climate concerns

Current policies fail to address the political and technological challenges for limiting global greenhouse-gas emissions, scientists warned on 13 July in a climate-risk report commissioned by the UK Foreign & Commonwealth Office. The combined efforts of the world's major economies, including China and the European Union, are unlikely to result in emissions being cut to safe levels, it says. It highlights the "enormous

COMING UP

18–22 JULY

About 1,200 researchers gather in Dresden, Germany, for the European Biophysics Congress. go.nature.com/r5uehz

18–23 JULY

Scientists and clinicians meet in Washington DC for the Alzheimer's Association International Conference. Topics include early detection, risk factors and clinical trials for Alzheimer's and dementia. go.nature.com/avsh5k

20 JULY

The crew of NEEMO (NASA Extreme Environment Mission Operations) is due to start its 14-day mission 19 metres under the sea, off Key Largo in Florida. NEEMO aims to prepare for future deep-space missions.

risks" to security if greater competition for land or water cause conflict. Climate-change risks need to be assessed in the same way as international security threats, the authors conclude.

Medical-cures bill

On 10 July, the US House of Representatives passed the 21st Century Cures Act, which would provide a US\$8.75-billion boost to the US National Institutes of Health over five years and speed regulatory approval for new antibiotics and medical devices. The bill also sets aside money for research into precision medicine and opportunities for young researchers. The bill will now need to be passed by the US Senate.

NATURE.COM

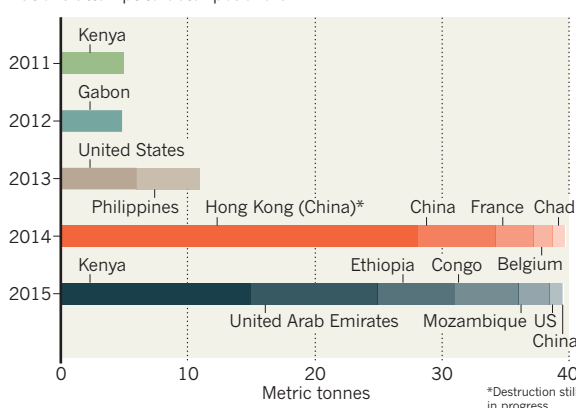
For daily news updates see:
www.nature.com/news

TREND WATCH

More ivory has already been destroyed this year in the fight against poaching than in any previous year. Last week, Mozambique burned 2,400 kilograms of seized elephant tusks and nearly 200 kilograms of rhino horn. Other nations, including the United States and China, have also burned or crushed ivory, in the hope of discouraging trafficking and poaching. High levels of poaching in Africa in 2014 are likely to be still driving down the number of elephants on the continent.

IVORY CRUSH

The amount of illegal ivory destroyed has escalated greatly as nations attempt to deter poachers.



NEWS IN FOCUS

BODY-ON-A-CHIP Drug companies put model organs to the test **p.266**

E-CIGARETTES Failure of California bill leaves regulatory hole **p.267**

PENTAQUARK Particle that was ruled out shows up at LHC **p.267**



EDUCATION Why dull science lectures are out, and active learning is in **p.272**

BEN SIMON/AFP/GETTY



A health worker vaccinates a child against polio in Kano, northern Nigeria, in 2013.

PUBLIC HEALTH

Smart shots bring Nigeria to brink of polio eradication

The nation has embraced the latest research and innovative approaches to vaccination.

BY EWEN CALLAWAY

Just three years ago, Nigeria was a threat to the global push to eradicate polio. Africa's most populous nation recorded 122 cases in 2012, more than all other countries combined, and funders of the eradication campaign were growing exasperated with the nation's faltering vaccination efforts and exportation of cases. Now Nigeria is on the brink of being free of the virus, thanks in large part to its embrace of innovative approaches to vaccination and public health.

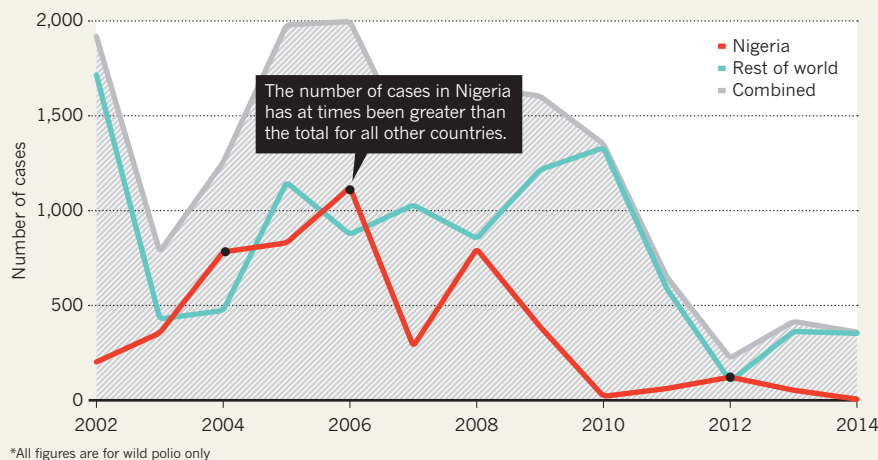
On 24 July, Nigeria will mark one year since it recorded a case of polio caused by the wild virus (mutated versions of live vaccine strains may still be circulating in poorly vaccinated areas but are considered easier to control than wild strains). If reached, this landmark will set the stage for polio's official eradication in Africa — which can only be declared after the region has been free of the virus for three consecutive years — and leave Pakistan and Afghanistan as the only remaining hosts of wild poliovirus.

A global eradication push began in 1988,

when 350,000 cases were recorded worldwide. By 1994, the Americas were polio free, and by the early 2000s Europe and most African countries had vanquished the virus (India followed much later, recording its last known case in 2011). But polio persisted in parts of Asia and in Africa, where the vast majority of cases were in northern Nigeria. Rumours that the vaccine caused sterility led to boycotts by several Nigerian states in 2003 and contributed to a bounce in cases (see 'Rise and fall of polio') that spread the virus back to many countries ►

RISE AND FALL OF POLIO

After cases started to increase in Nigeria in the early 2000s, the country exported the virus around the world; a smart approach to public health and vaccination has turned things around.



► in Africa, and as far afield as Indonesia.

Complacency at all levels of the Nigerian government meant that vaccine campaigns were often incomplete. Worsening security in the country's north-east — large parts of which were, and still are, under the control of the Islamic militant group Boko Haram — made matters worse. "If you want to know anything about polio not going right, this is the place to come," said a veteran of India's successful eradication campaign back in December 2012, while working on the Nigerian campaign.

Eventually, pressure from other African countries convinced Nigeria's leaders to take polio seriously, says Heidi Larson, an anthropologist at the London School of Hygiene & Tropical Medicine who studies vaccination. Meanwhile, as they saw children continue to be crippled by the disease, religious and traditional leaders in the Muslim-dominated north started to support, rather than oppose, polio eradication.

TREND SETTER

In 2012, Nigeria created a national emergency operations centre to coordinate anti-polio efforts between government and international organizations, such as the World Health Organization (WHO); similar hubs followed in key northern states. Efforts turned to identifying and mapping communities that had been routinely missed in vaccination campaigns,

such as nomadic groups, people living near state and local borders and those who turned vaccinators away. "You probably have the best maps of all of Africa in northern Nigeria," says Michael Galway, who heads polio eradication efforts at the Bill & Melinda Gates Foundation in Seattle, Washington.

Vaccination teams carried detailed 'micro-plans' that identified each house to visit and global-positioning units that meant their supervisors could ensure the plans were followed. In areas in the north-east that were controlled by Boko Haram, health officials mounted 'hit and run' campaigns, rapidly vaccinating children on short notice when security conditions allowed.

Daniel Santong, a Nigerian field epidemiologist who is part of a US Centers for Disease Control and Prevention effort to map nomadic communities, says that most of the Fulani people he tracks welcome polio vaccination of children, and that more-sceptical communities have been won over with health camps that provide care and medicines to adults in remote areas. "This attracts them to bring their children to come to immunization," he says.

Nigeria's success in battling polio has been aided by its embrace of vaccine research, notes Nicholas Grassly, an epidemiologist at Imperial College London. When clinical results showed that an oral vaccine against the two types of

poliovirus that were still circulating in the wild was more effective than a vaccine against all three strains¹, Nigerian campaigns quickly adopted it. And last year, the country delivered injectable polio vaccines to more than 3 million children in high-risk areas after studies showed that children who had previously received an oral version received an added immune boost from the jab^{2,3}. Grassly says that injectable vaccines will be especially important in eradicating the last pockets of vaccine-derived polioviruses.

Assuming that no new polio cases are recorded after 24 July, samples from around Nigeria will be analysed for traces of the disease. If no sign is found, Nigeria will be removed from the list of polio-endemic countries, probably in late August or early September, says Hamid Jafari, the director of the Global Polio Eradication Initiative at the WHO. Africa could then be declared polio-free as soon as August 2017 — 3 years after the continent's last known case, in Somalia, in August 2014, which was caused by a virus imported from Nigeria.

Attention is then likely to shift to Pakistan and Afghanistan, where a combined total of 29 polio cases have been reported so far this year; a vast improvement from a year ago. Twenty-five of those cases were in Pakistan, which recently opened an Emergency Operations Centre that was modelled on Nigeria's, Galway notes. Nigeria's success in vaccinating children in Boko Haram-controlled areas suggests that the security challenges in Afghanistan and Pakistan are not insurmountable, says Larson. "I think that polio will go on to be eradicated," adds David Heymann, chair of Public Health England and former head of the WHO's polio eradication programme.

But Nigeria must guard against complacency, says Jafari. "We clearly need more and more surveillance to be comfortable that transmission has been interrupted," he says, pointing out that remote areas of neighbouring Chad, Niger and Cameroon have not received the same level of attention as Nigeria in terms of surveillance for the virus and vaccination. Galway agrees: "The biggest risk we face is falling back and saying we're finished when we're not." ■

1. Sutter, R. W. *et al. Lancet* **376**, 1682–1688 (2010).
2. Jafari, H. *et al. Science* **345**, 922–925 (2014).
3. John, J. *et al. Lancet* **384**, 1505–1512 (2014).

SOURCE: GLOBAL POLIO ERADICATION INITIATIVE


**MORE
ONLINE**

TOP STORY



Energetic bacteria form frictionless superfluids
go.nature.com/plr93z

MORE NEWS

- Astronomers spy brightest-ever supernova go.nature.com/yvlnco
- Speedy study claims climate change doubled chances of European heatwave go.nature.com/w6u92r
- Oldest animal sperm found in worm cocoon fossil go.nature.com/1ylzmc

NATURE PODCAST



Organic molecules in space, treating brain injuries and training children to think like scientists.
nature.com/nature/podcast

VISUALS UNLIMITED/CORBIS

CLINICAL TRIALS

Alzheimer's data lawsuit is sign of growing tensions

Battle between California universities raises questions about research ownership.

BY ERIKA CHECK HAYDEN

In a case that highlights growing tensions over who controls scientific data after faculty members move on, a California university has sued a researcher who left for another institution in June, alleging that he and colleagues conspired to hijack a valuable trove of Alzheimer's data. The researcher insists that he acted properly.

On 2 July, the University of California, San Diego (UCSD), filed a lawsuit alleging that when Alzheimer's researcher Paul Aisen took a new job at the University of Southern California (USC), he improperly tried to take with him a large clinical-trials programme that he was leading.

Aisen ran the Alzheimer's Disease Cooperative Study (ADCS), a 24-year-old clinical-trials network, at UCSD from 2007 until 21 June, when he left for USC. Within two weeks, UCSD sued Aisen, colleagues who left with him, and USC, alleging a conspiracy to transfer contracts and data related to the ADCS. The researchers are also accused of blocking their colleagues back at UCSD from "administrative control" of the ADCS database.

DATA DISPUTE

UCSD says that Aisen's alleged attempt to transfer the study and data without UCSD's permission is neither legal nor usual. "When a researcher leaves a university, there is always proper notification and a transition plan. None of which occurred here. Never does a researcher simply take control of data without permission from the university, the grant sponsors and the [US National Institutes of Health]," says UCSD health-sciences spokesperson Jacqueline Carr.

Aisen counters that he left UCSD after what he saw as long-standing difficulties with running the ADCS there. "I did not feel that our research had the necessary support at UCSD," he told *Nature*. He says that he did nothing out of the ordinary when he left UCSD, that he remains co-principal investigator on several clinical trials that are coordinated by the ADCS, and that UCSD still has access to the study network's data.

Aisen adds that several times before and after his departure, he asked UCSD to allow



The campus of University of California, San Diego, which hosted a major Alzheimer's project.

him to transfer the ADCS programme, but that these requests were rebuffed. "The reason given was, this programme will always be at UCSD, period," Aisen told *Nature*.

Legal disputes over the ownership of data are not unheard of. But such arguments are escalating as universities become more involved in the development of treatments, a trend driven by diminishing funding by pharmaceutical companies for work to bridge the 'valley of death' between basic research and clinical trials.

The ADCS is not a single study, but a network that coordinates 11 clinical trials in Alzheimer's disease, involving researchers at 70 institutions in the United States and Canada. The US National Institute on Aging in 2013 awarded UCSD and Aisen up to US\$55 million to run the ADCS until 2018. Two companies, Eli Lilly of Indianapolis, Indiana, and the Japan-based Toyama Chemical, currently sponsor clinical trials coordinated by the ADCS. UCSD estimates in its lawsuit that the total funding for studies under the ADCS's umbrella is more than \$100 million.

The university also alleges that Aisen began discussing moving to USC in April, that he discussed a transfer of the Alzheimer's study to USC, and that his staff discussed with

Toyama the possibility of transferring its contract from UCSD to USC.

The complaint further alleges that Aisen notified UCSD by e-mail on 18 June of his intent to depart for USC on 1 July, and that in-house lawyers for UCSD then advised Aisen not to take with him "equipment, records, electronic data or software that were purchased or created for the ADCS at UC San Diego". Then, says the lawsuit, Aisen sent another e-mail on 21 June declaring his immediate resignation. Some of the Alzheimer's cooperative study staff changed administrative access codes and passwords for databases maintained through the study, the complaint alleges, and then abruptly resigned themselves.

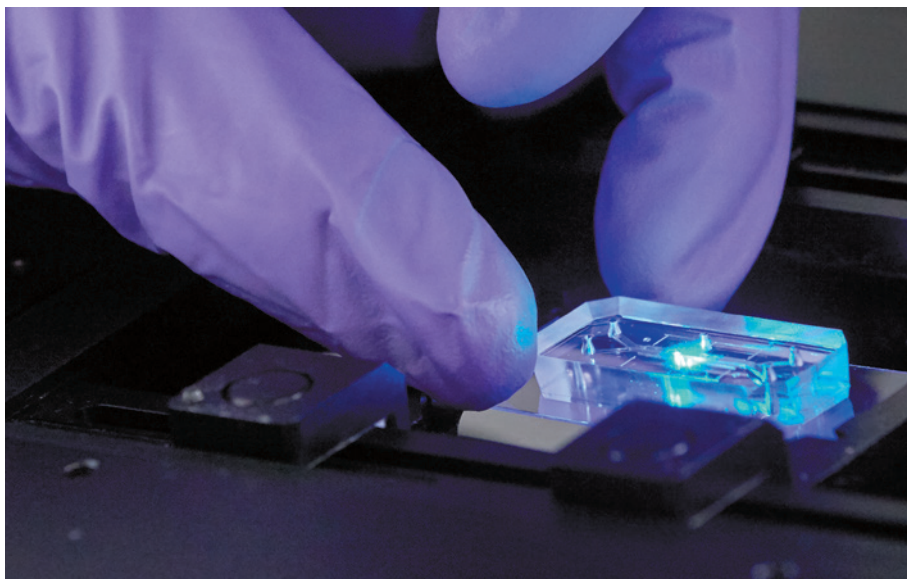
Aisen says that his staff retained administrative control over the ADCS database to maintain the integrity of highly regulated clinical trial data. He denies that his staff changed passwords. "That claim is not true," he told *Nature*.

FUNDING WOES

In a court filing, Aisen states that he left UCSD because of "funding delays and shortfalls that I believe hampered the work of ADCS". For instance, he says that he had too few staff to help the ADCS to negotiate contracts and salary restrictions, which "impacted ADCS's hiring abilities". The court has received letters supporting Aisen's oversight of individual studies and of ADCS data from colleagues at Harvard Medical School in Boston, Massachusetts; Yale University School of Medicine in New Haven, Connecticut; Wake Forest School of Medicine in Winston-Salem, North Carolina; and from Lilly.

According to UCSD's lawsuit, USC offered Aisen a \$500,000 annual salary until 2020, as well as low- and no-interest home loans that would be partially forgiven over time. Aisen told *Nature* that salary and compensation were never part of his discussion with USC.

"I was surprised that Paul Aisen abruptly decided to take a better offer, but I understand it perfectly," says Jeanne Loring, a neuroscientist and stem-cell researcher at the Scripps Research Institute in La Jolla, California. "Funding is so uncertain; having the opportunity to follow through on what you've been working on in science is attractive. I would have done the same thing." ■



Miniature devices that mimic human organs could help to replace animals used in drug testing.

BIOTECHNOLOGY

'Organs-on-chips' go mainstream

Drug companies put in vitro systems through their paces.

BY SARA REARDON

Researchers who are developing miniature models of human organs on plastic chips have touted the nascent technology as a way to replace animal models. Although that goal is still far off, it is starting to come into focus as large pharmaceutical companies begin using these *in vitro* systems in drug development.

"We are pretty excited about the interest we get from pharma," says Paul Vulto, co-founder of the biotechnology company Mimetas in Leiden, the Netherlands. "It's much quicker than I'd expected." His company is currently working with a consortium of three large pharmaceutical companies that are testing drugs on Mimetas's kidney-on-a-chip. At the Organ-on-a-Chip World Congress in Boston, Massachusetts, last week, Mimetas was one among many drug and biotechnology firms and academic researchers showing off the latest advances in miniature model organs that respond to drugs and diseases in the same way that human organs such as heart and liver do.

"We're surprised at how rapidly the technology has come along," says Dashyant Dhanak, global head of discovery sciences at Johnson & Johnson in New Jersey, which

announced last month that it would use a thrombosis-on-chip model from Massachusetts biotechnology firm Emulate to test whether experimental and already-approved drugs could cause blood clots.

Proponents of organs-on-chips say that they are more realistic models of the human body than are flat layers of cells grown in Petri dishes, and could also be more useful than animal models for drug discovery and testing. A lung-on-a-chip, for instance, might consist of a layer of cells exposed to a blood-like medium on one side and air on the other, hooked up to a machine that stretches and compresses the tissue to mimic breathing.

But while some companies are developing chips to mimic diseased organs, most are still testing whether existing drugs behave in the chips as they do in healthy human tissues. And chips or not, any new drug must first be tested in healthy humans for safety, says James Hickman, a bioengineer at the University of Central Florida in Orlando. Using an *in vitro* organ might help to eliminate or shorten this step.

"We are surprised at how rapidly the technology has come along."

The chips could also help companies to pinpoint the dose of a drug that is both effective and safe, says Matthew Wagoner, a drug-safety scientist at AstraZeneca in Waltham, Massachusetts. If regulators accept such data, the method might eventually allow companies to skip the portion of a clinical trial that tests a wide range of drug doses on patients.

Other researchers are eager to use organs-on-chips to illuminate differences between animal models and humans. Adrian Roth, head of *in vitro* safety research at Roche in Basel, Switzerland, says that such comparisons were useful when one of Roche's experimental drugs was found to cause liver tumours in rats. Roche used data from *in vitro* models of human and rat livers to argue that the mechanism causing the liver tumours was unique to rodents and should not prevent studies in humans.

But some users are concerned that the hype is getting ahead of reality. The danger, Roth says, is that companies will abandon the technology if it fails to live up to inflated expectations, as has happened with the use of genomics for personalized medicine. "As a pharma company, you have to be very pragmatic," he says, and not expect the chips to replace animals all at once.

There is reason to be hesitant. Many presenters at the Boston meeting validated their model organs by showing that they responded to drugs in the same way that a human organ does. A model heart might be tested to see whether it speeds up after a dose of adrenaline, for instance. But such tests do not capture anything like the full complexity of organ function, and chips may struggle to recreate aspects of functioning that are governed by complex signals from, say, the endocrine and immune systems.

Even testing a known drug in a system that hitches multiple organ chips together might be difficult to validate, Roth adds, because researchers might not know what to look for. The potentially toxic effect of the painkiller paracetamol on the liver, for instance, is well characterized, but less is known about how other organs respond to the drug.

Nevertheless, many pharmaceutical companies say that organs-on-chips are now sufficiently advanced to justify investment in their use and refinement. "We think it's important to be involved," says Michelle Browner, senior director of platform innovation at Johnson & Johnson. Only that way can the technology be developed in line with what the company needs, she says.

And government regulators are also interested. This autumn, the US National Center on Advancing Translational Sciences (NCATS) in Bethesda, Maryland, will bring together academic scientists, pharmaceutical companies and regulators to discuss the chips' use. NCATS is also funding 11 research teams, each of which is developing a different organ or system that will eventually be linked together to build an entire 'body-on-a-chip'. ■

EMULATE

POLITICS

Anti-vaping bill goes up in smoke

Demise of California legislation highlights rise of lobbying.

BY DANIEL CRESSEY

The failure last week of an attempt to extend California's smoking ban to electronic cigarettes has disappointed researchers who worry that a boom in vaping will re-normalize smoking in places where it now carries social stigma, and lead to a new generation of people addicted to nicotine. Although several other US states have already passed e-cigarette legislation, the California bill had special resonance because the state has a reputation for pioneering anti-tobacco legislation.

The bill's demise, says Vaughan Rees, director of the Center for Global Tobacco Control at the Harvard T. H. Chan School of Public Health in Boston, Massachusetts, also highlights the rise of intensive political lobbying in this arena. "The e-cigarette industry is acquiring the kind of influence that the conventional tobacco industry has used for many years to prevent the implementation of sensible measures to protect the health of the public," he says.

Since the modern e-cigarette emerged a decade ago, the United States has taken to the

technology in a big way (see *Nature* 513, 24–26; 2014). In 2013, around 2.5% of people across the country reported having used the devices in the past month; among adolescents, e-cigarettes are now used more than conventional cigarettes, said the US Centers for Disease Control and Prevention in April this year.

But regulation has not kept up. The Food and Drug Administration (FDA) proposed a rule for regulating the devices in April 2014, but the rule has been mired in political controversy ever since, leaving an absence of country-wide laws. Even if the FDA does regulate e-cigarettes, it will still be up to local law-makers to decide where the devices can be used, and how much they will be taxed. North Dakota, New Jersey, Utah and Hawaii prohibit use of e-cigarettes in public spaces, and nearly 400 municipalities have similar laws, says the American Cancer Society Cancer Action Network in Washington DC, which campaigns for cancer-fighting legislation.

In January, the California Department of Public Health asked physicians to urge e-cigarette users to quit, and released a report warning that without more regulation, "it is

likely that California's more than two decades of progress to prevent and reduce traditional tobacco use will erode".

The California bill, pushed by state senator Mark Leno (Democrat), would have regulated e-cigarettes in the same way as conventional tobacco products, making it illegal to use them in restaurants, bars, hospitals and workplaces. Leno withdrew the bill on 8 July, after a committee in the state assembly amended it to such an extent that he said it had become pointless.

Whether to tax e-cigarettes at the same rate as tobacco is also a contentious issue. Advocates of e-cigarettes say that vaping is less dangerous than smoking, and that taxing e-cigarettes at below the rate for tobacco makes it more attractive. "Any time they are vaping, they're not smoking a cigarette. That's a win," says Cynthia Cabrera, executive director of the Smoke-Free Alternatives Trade Association in Washington DC. Critics contend that lower taxation encourages people to maintain their nicotine addictions, and to start using the drug in the first place, perhaps as a gateway to tobacco. Washington DC will put a 70% mark-up on the devices in October.

Researchers are broadly split into those who think that the products undermine attempts to free society from nicotine addiction entirely, and thus should be subject to the same restrictions and taxes as tobacco, and those who think that e-cigarettes deserve lighter regulation than normal cigarettes because they could help smokers to quit. Still, even proponents of the harm-reduction strategy would support reasonable efforts to limit marketing and sale of e-cigarettes to minors, says Rees. ■

PARTICLE PHYSICS

Forsaken pentaquark particle spotted at CERN

Exotic subatomic species confirmed at Large Hadron Collider after earlier false sightings.

BY MATTHEW CHALMERS

An exotic particle made up of five quarks has been found a decade after experiments seemed to rule out its existence.

The short-lived 'pentaquark' was spotted by researchers analysing data on the decay of unstable particles in the LHCb experiment at the Large Hadron Collider (LHC) at CERN, Europe's particle-physics laboratory near Geneva. The finding, says LHCb spokesperson Guy Wilkinson, opens a new era in physicists' understanding of the strong nuclear

force that holds atomic nuclei together.

"The pentaquark is not just any new particle — it represents a way to aggregate quarks, namely the fundamental constituents of ordinary protons and neutrons, in a pattern that has never been observed before," he says. "Studying its properties may allow us to understand better how ordinary matter, the protons and neutrons from which we're all made, is constituted."

Protons and neutrons are made up of three kinds of quarks bound together, but theorists calculate that, in principle, particles could be made of up to five quarks. Such particles would

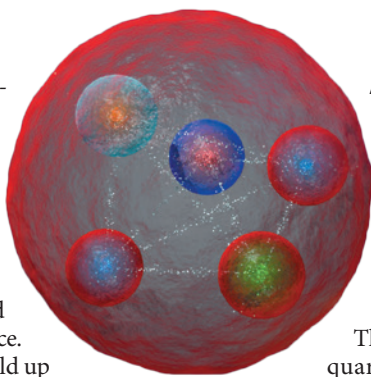
be rich testing grounds for quantum chromodynamics (QCD) — the theory that describes the forces that hold quarks together.

In 2002, researchers at the SPring-8 synchrotron in Harima, Japan, caused a stir when they announced that they had discovered a pentaquark, roughly 1.5 times heavier than a proton, inferring its existence from the debris of collisions between high-energy photons and neutrons. Within a year, more than ten other labs had reported finding evidence for the particle by reanalysing data. But many others saw no evidence for such a state and, in 2005, ►

► the discovery was pronounced a miracle. The final straw came with an experiment at the Thomas Jefferson National Accelerator Facility in Newport News, Virginia, that repeated the SPing-8 measurement with more data and ruled out the pentaquark's existence.

The episode has been held up as an example of how scientists can be tricked by data into seeing more than is there. In 2008, the annual *Review of Particle Physics*¹ described the pentaquark as “a curious episode in the history of science”, and its recent listings have no dedicated entry for the particle.

But the LHCb result leaves little doubt that pentaquarks are real, researchers say. Physicists saw a signal showing the unexpected appearance of two short-lived objects weighing 4.38 and 4.45 gigaelectronvolts (4.67 and 4.74 times heavier than a proton) during the decay of trillions of subatomic particles known as ‘Lambda B’ baryons, by analysing data from 2009–12.



Artist's impression of a pentaquark.

After exhausting other known particles as candidates, the team concluded that the new objects correspond to a pentaquark in two different configurations.

The particle contains two ‘up’ quarks, a ‘down’ quark, and a ‘charm’ quark–antiquark pair, making it a ‘charmonium’ pentaquark. A preprint on the find was posted on the arXiv server on 14 July², and has been submitted for publication in *Physics Review Letters*. “It’s about the most exciting discovery in QCD I could imagine,” says Frank Wilczek, one of the original architects of QCD, at the Massachusetts Institute of Technology.

The result, which first caught the attention of physicists on LHCb in 2012 as a bump in their data, was a total surprise, says LHCb’s Sheldon Stone, at Syracuse University in New York. “In the old days we searched for new particles by bump-hunting, but in this case the bump found

us!” he says. “For historical reasons we were quite haunted by the word pentaquark, so we did every conceivable check we could,” he says. The LHCb team says that there is a vanishingly small chance of the signal appearing if no new particles existed. Their statistical bar — known as 9-sigma — is higher than the 5-sigma usually required for a discovery in particle physics.

“If I have an immediate feeling of worry it is that they claim two states: is this because they have found a process that favours production of pentaquarks, or because they have not really found the best interpretation of the data?” says theorist Frank Close, of the University of Oxford, UK.

The new pentaquark is not the one, known as the theta+, seen back in 2002: it is almost three times heavier, and contains different kinds of quarks. “I think our result will energize the search for many different pentaquark states, including the debunked theta+,” says Stone. ■

1. Particle Data Group. *Phys. Lett. B* **667**, 1061–1206 (2008).
2. LHCb collaboration. Preprint at <http://arxiv.org/abs/1507.03414> (2015).

PSYCHIATRY

First robust genetic links to depression emerge

Discoveries energize hunt for genes connected to mental illness.

BY HEIDI LEDFORD

No one was more surprised than Jonathan Flint when his project — an effort to find genetic sequences linked to depression — showed the first hint of success 18 months ago. He knew the odds were slim: a study of 9,000 people with major depressive disorder had come up empty¹, and Flint had heard rumours that a follow-up analysis of 17,000 people had also met with disappointment. “I thought, ‘There’s no way,’” says the geneticist from the University of Oxford, UK, whose study had by that point analysed only 5,303 people with depression.

Flint has proved himself wrong. In *Nature* this week, his team reports² the first two genetic markers reproducibly linked to major depressive disorder, one of the leading causes of disability globally. The findings could guide biologists to new drugs, and could one day be used to aid diagnosis. But many in the field are excited that the markers have been unearthed at all. The results look set to end years of debate over whether sequences for such a complex

disorder could be found — and Flint’s study may serve as a framework for future attempts to collect data from tens of thousands of people.

More than 350 million people have depression. The disorder’s symptoms and severity can vary widely from one person to the next, and particularly between men and women. This suggests that different conditions have been lumped together into one diagnosis, complicating genetic analyses (see *Nature* **515**, 182–184; 2014).

The few hits from early studies attempting to find genetic sequences linked to depression had disappeared on closer scrutiny, so Flint knew that he would need samples from thousands of people, and a way to reduce the variability in their illness. Flint and Kenneth Kindler, a psychiatrist at Virginia Commonwealth University in Richmond who is renowned for his diagnostic prowess, decided to do the study in China, because of its large population and because depression is believed to be under-diagnosed there. In that climate, Flint reasoned, those who are diagnosed are likely to share a severe form of the disorder. To reduce the variability further,

his team also limited the study to women of Han Chinese ethnicity.

By early 2014, Flint, Kindler and a team of collaborators had analysed DNA sequences from 5,303 Chinese women with depression, and another 5,337 controls. As Flint expected, 85% of the depressed women had a severe form of the disorder called melancholia, which robs people of the ability to feel joy. “You can be a dotting grandparent and your favourite grandchildren can show up at your door,” says Douglas Levinson, a psychiatrist at Stanford University in California, “and you can’t feel anything.”

The analysis yielded two genetic sequences that seemed to be linked to depression: one in a stretch of DNA that codes for an enzyme whose function is not fully understood, and the other next to the gene *SIRT1*, which is important for energy-producing cell structures called mitochondria. The correlations were confirmed in another set of more than 3,000 depressed men and women and over 3,000 controls.

The mitochondrial connection chimes with previous work, including some from Flint’s lab³, that has linked mitochondrial abnormalities to



Depression can take many forms, making it hard to tease out genetic links.

depression. “It’s an appealing bit of biology for a disorder that makes people tired and unmotivated,” says Levinson.

But he is most intrigued by the way Flint and his colleagues designed their study. Levinson says that the Psychiatric Genomics Consortium, of which he is a member, has analysed data from 17,000 people with depression without finding a genetic hit, as Flint had heard. In light of Flint’s results, the group, which is made up of researchers who agree to pool genetic data, is now investigating whether limiting the analysis to people with particularly severe depression

might change things.

If it does, still more samples will be needed before other genetic links emerge. Levinson expects a series of biobanking efforts in places including the United Kingdom, the Netherlands and Australia to provide tens of thousands of genomes for analysis within the next five years. The hope is that as more genetic links are found, they will flag up groups of proteins known to work together to affect certain cellular functions: these ‘pathways’ could be investigated as drug targets, and for their potential to make diagnosis of depression more definitive.

Flint’s success may energize that search, says Patrick Sullivan, a psychiatric geneticist at the University of North Carolina at Chapel Hill. “We’ve had to learn not to listen a lot to our critics,” he says. “If we listened to people telling us that what we’re doing is stupid, we would have stopped years ago.” ■

1. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium *Mol. Psychiatry* **18**, 497–511 (2013).
2. CONVERGE Consortium *Nature* <http://dx.doi.org/10.1038/nature14659> (2015).
3. Cai, N. *et al. Curr. Biol.* **25**, 1146–1156 (2015).

CORRECTIONS

The News Feature ‘Weighing the world’s trees’ (*Nature* **523**, 20–22; 2015) incorrectly described the photosynthesis measurements made by the Orbiting Carbon Observatory-2. Those observations are made by the same spectrometers that measure CO₂ concentrations.

The News story ‘Researchers pin down risks of low-dose radiation’ (*Nature* **523**, 17–18; 2015) incorrectly calculated an ‘expected’ death rate from leukaemia among the workers. Only 30 deaths were attributable to the radiation, and the relative risk increment posed by each additional 10 mSv was 3%.



THE SCIENTIST OF THE FUTURE

A special issue of Nature examines what is needed to grow the next generation of scientists.



What does it take to be a successful scientist in the modern world? The obvious answers are deep knowledge of a discipline and mastery of the scientific method. But there are other key requirements, such as the ability to think critically and solve problems creatively and collaboratively. Communication skills are a must, and mastery of modern technology helps.

ILLUSTRATION BY VASAVA

For generations, classes in science, technology, engineering and maths (STEM) have been focused almost exclusively on building knowledge alone. A steady diet of lecture-based learning was designed to fill students up with facts and test their ability to memorize them. Teaching the other skills was too often given short shrift.

Now educators and education researchers are calling for change. They argue that creative thinking, problem solving, motivation, persistence and other 'twenty-first-century skills' can, and should, be taught and fostered through well-designed courses. Developing these skills enhances students' abilities to master and retain knowledge; many hope that focusing on them will help to curb the alarming rate at which students interested in STEM abandon the subjects. The Organisation for Economic Co-operation and Development deems STEM education as crucial to powering innovation and economic growth, and has strongly encouraged investment in education strategies that focus on twenty-first-century skills.

Now *Nature*, in collaboration with *Scientific American*, is taking a look at the challenges in STEM education (a full listing of content is available at nature.com/stem). A News Feature on page 272 discusses the move towards 'active learning' rather than passive lecturing in the undergraduate classroom, but finds that encouraging innovative methods requires a change in incentives. A Comment article by representatives of the Association of American Universities and the Research Corporation for Science Advancement Cottrell Scholars on page 282 offers a road map for the institutional changes that will be required to shift the status quo.

Those teaching science in primary and secondary schools face different constraints, but have no shortage of innovative practices. A News Feature on page 276 looks at some of the most creative STEM education programmes around the world, for preschoolers up to teens. On page 286, leading design practitioners explain how nature itself aids early child development, and how architecture and play spaces are best engineered for learning. At the other end of the spectrum, senior researchers should brush up their leadership skills, says a Comment piece on page 279.

Finally, *Nature* polled some of the leading thinkers in science and education for what it takes to make an effective scientist in the twenty-first century. With answers on page 371 that range from the practical to the philosophical, it is clear that the science classroom is in for a radical change. ■



THE SCIENCE OF TEACHING SCIENCE

Active problem-solving confers a deeper understanding of science than does a standard lecture. But some university lecturers are reluctant to change tack.

BY M. MITCHELL WALDROP



Outbreak alert: six students at the Chicago State Polytechnic University in Illinois have been hospitalized with severe vomiting, diarrhoea and stomach pain, as well as wheezing and difficulty in breathing. Some are in a critical condition. And the university's health centre is fielding dozens of calls from students with similar symptoms.

This was the scenario that 17 third- and fourth-year undergraduates dealt with as part of an innovative virology course led by biologist Tammy Tobin at Susquehanna University in Selinsgrove, Pennsylvania. The students took on the role of federal public-health officials, and were tasked with identifying the pathogen, tracking how it spreads and figuring out how to contain and treat it — all by the end of the semester.

Although the Chicago school and the cases were fictitious, says Tobin, “we tried to make it as real as possible”. If students decided to run a blood test or genetic assay, Tobin would give them results consistent with enterovirus D68, a real respiratory virus. (To keep the students from just getting the answer from the Internet, she portrayed the virus as an emergent strain with previously unreported symptoms.) If they decided to send a team to Chicago, Tobin would make them look at real flight schedules and confirm that there were enough seats.



THE 21ST CENTURY SCIENTIST
A *Nature* and *Scientific American*
special issue nature.com/stem

In the end, the students pinpointed the virus, but they also made mistakes: six people died, for example, in part because the students did not pay enough attention to treatment. However, says Tobin, “that doesn’t affect their grade so long as they present what they did, how it worked or didn’t work, and how they’d do it differently”. What matters is that the students got totally wrapped up in the problem, remembered what they learned and got a handle on a range of disciplines. “We looked at the intersection of politics, sociology, biology, even some economics,” she says.

Tobin’s approach is just one of a diverse range of methods that have been sweeping through the world’s undergraduate science classes. Some are complex, immersive exercises similar to Tobin’s. But there are also team-based exercises on smaller problems, as well as simple, carefully tailored questions that students in a crowded lecture hall might respond to through hand-held ‘clicker’ devices. What the methods share is an outcome confirmed in hundreds of empirical studies: students gain a much deeper understanding of science when they actively grapple with questions than when they passively listen to answers.

“We find up to 20% better grades over usual methods,” says Tom Duff, a computer scientist who developed a team-based learning approach at the University of the West of Scotland in Paisley, UK. Other active-learning proponents have found similar gains. Last year, a group led by biologist Scott Freeman at the University of Washington in Seattle published an analysis of 225 studies of active learning in science, technology, engineering and mathematics (STEM) and found that active learning cut course failure rates by around one-third¹.

“At this point it is unethical to teach any other way,” declares Clarissa Dirks, a microbiologist at the Evergreen State College in Olympia, Washington, and co-chair of the US National Academies Scientific Teaching Alliance, an initiative to reform undergraduate STEM education.

Active learning is winning support from university administrators, who are facing demands for accountability: students and parents want to know why they should pay soaring tuition rates when so many lectures are now freely available online. It has also earned the attention of foundations, funding agencies and scientific societies, which see it as a way to patch the leaky pipeline for science students. In the United States, which keeps the most detailed statistics on this phenomenon, about 60% of students who enrol in a STEM field switch to a non-STEM field or drop out² (see ‘A persistence problem’). That figure is roughly 80% for those from minority groups and for women.

TOUGH SELL

Not everyone embraces the idea. Active learning can be a tough sell to faculty members who thrived on standard lectures during their own student years, and who wonder whether the benefits of active learning — which requires substantially more preparation than do standard lectures — could possibly justify the time that the approach would take away from their research.

Understanding and addressing the resistance has become one of the reformers’ prime concerns. Robert Lue, the other co-chair of the teaching alliance and director of the Derek Bok Center for Teaching and Learning at Harvard University in Cambridge, Massachusetts, says that he is “hell bent on erasing this sense that research is where you apply your intellect, and teaching is a rote skill”. Scientists need to approach teaching with the same rigour and appreciation for evidence that they exercise in the laboratory, he says. “It’s at the frontier of research. And the more people we get involved, the faster that research will go.”

On the surface, active-learning classes can seem to differ little from more conventional approaches. Undergraduate students have always had discussion sessions to ask about the course material, and laboratory classes in which they would carry out experiments. But if you look more closely, says Tobin, these are often just ‘cookbook’ exercises. The typical approach is ‘read that and be prepared to talk about these questions’, or ‘follow that procedure and you’ll get this result’. In an active-learning class such as hers, she says, the students take charge

of their own education. “They are framing the questions themselves.”

The same is true for active learning in first-year courses, in which the teachers often do supply the questions — but frame them in a way that asks for more than a rote recitation of facts. It is the difference between

“AT THIS POINT IT IS UNETHICAL TO TEACH ANY OTHER WAY.”

‘name the sensory nerves of the leg’, and what neuroscientist Sarah Leupen asks of her introductory physiology class at the University of Maryland, Baltimore County (UMBC):

You’re innocently walking down the street when aliens zap away the sensory neurons in your legs. What happens?

- a) Your walking movements show no significant change.
- b) You can no longer walk.
- c) You can walk, but the pace changes.
- d) You can walk, but clumsily.

“We usually get lots of vigorous debate on this one,” says Leupen, who spends most of her class time firing such questions at her students. “It’s lovely to experience.”

What makes those questions special is that the students cannot answer them simply by reading the course material — although they are expected to have done that before attending class. Instead, they have to apply what they have learned, which they do by clustering around tables in small teams and arguing over the options. That struggle is the real pay-off, says Leupen, who eventually explains the right answer (in this case, d). And if a team gets it wrong, she says, “that’s usually a good thing — because then they really remember it”.

Evidence has been accumulating for decades that students who actively engage with course material will end up retaining it for much longer than they would have otherwise, and they will be better able to apply their knowledge broadly. But the evidence began to draw widespread attention only around the turn of the century — not least thanks to Carl Wieman, who suddenly became one of the movement’s most visible champions when he was awarded the 2001 Nobel Prize in Physics for his co-discovery of Bose–Einstein condensates. “I started way before the Nobel prize,” says Wieman, who is now at Stanford University in California. “It’s just that people didn’t pay attention to me until then.”

Wieman’s conversion began in the late 1980s, when he noticed something about the graduate students coming into his atomic-physics lab — then at the University of Colorado Boulder. “They had done really well as undergraduates, but couldn’t do research,” he says. Over the years, they learned how to be good scientists, “but that had little to do with how well they had done in their courses”.

In trying to figure out why, Wieman came across the already huge body of empirical research on learning — most of it totally unknown to science departments. Among the most striking findings, he says, was one³ that explained his own observation. It showed that in the traditional way of teaching, students could pass the test, but did not get a basic conceptual model of the subject, he says.

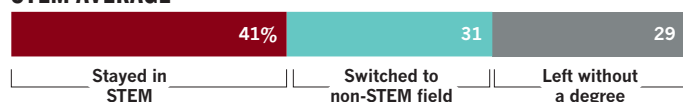
Other scientists were coming to much the same realization, and they were starting to experiment with other ways to teach. By the time Wieman’s Nobel shone a spotlight on the efforts, many fields had started what is now known as Discipline-Based Education Research: investigations into active methods for teaching concepts specific to each branch of science⁴.

Other powerful advocates included biologist Bruce Alberts, then president of the US National Academy of Sciences. In 2004, Alberts consolidated several academy panels into the Board on Science

A PERSISTENCE PROBLEM

A study tracking 17,000 post-secondary students in the United States and Puerto Rico found that only two-fifths of those who enrolled in a STEM discipline went on to obtain a degree in the field, or were still studying for one 6 years later.

STEM AVERAGE



ENGINEERING



PHYSICAL SCIENCES



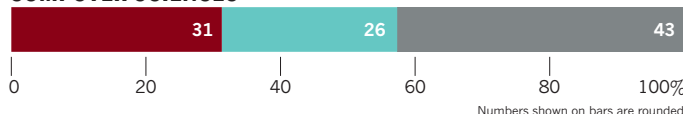
LIFE SCIENCES



MATHS



COMPUTER SCIENCES



Education: a group of senior scientists, initially chaired by Wieman, that has gone on to release a series of reports on education reform. The most recent of those, published in January 2015, is essentially a how-to manual for applying active learning in undergraduate settings⁵.

CULTURE SHOCK

Yet there is still plenty of scepticism, says Linda Hodges, a biochemist and head of the Faculty Development Center at the UMBC, and author of a forthcoming book on overcoming obstacles to education reform. One big reason, she says, is that for many scientists, active learning is sharply at odds with their beliefs about teaching.

Researchers often feel that a teacher's job is simply to communicate content: the factual knowledge covered in the course. That is a big stumbling block for active learning, because time spent on team discussions and the like can seem like time taken away from that content. Getting past that requires compromise, says Jeff Leips, a geneticist who teaches ecology and evolution at the UMBC. "You have to accept that you can't cover everything to the same level." But the pay-off is that the students retain much more of the material that is covered, and are able to use that knowledge much more effectively.

Another common belief, says Hodges, is that it is the professor who should be in control of the classroom. "A lot of these pedagogies ask you to relinquish some of your control and hand it over to a bunch of novice students," she says. "And that sounds odd to a lot of faculty." It definitely requires a different set of skills, agrees Leips. "A lecture is like a performance: you know the script." But an active-learning class is more like improvisational acting, he says. "You have to go with the flow", responding to questions and situations as they arise. And not everyone is comfortable with that.

Adding to the resistance is that many faculty members who try active learning hastily back off when the techniques do not seem to work, or when students start to turn in teacher evaluations that say, 'I had to teach myself!' or, 'Just tell me what I need to know!'.

One faculty member, who asked not to be named so that she could speak freely about her institution, tells the story of a chemistry instructor who told his students to 'work together', and then spent the rest of the class time reading. "Active learning done badly is worse

than a good lecture," says Leips. A 2011 survey of biology teachers at 77 US universities found that, even though most of them claimed to be using at least some active-learning methods, few of them were doing it properly⁶.

Proponents of active learning say that change will come only when innovations are made at every level of a university system (see page 282). To help interested biology teachers to do better, the US National Academies has been running a series of five-day workshops every summer since 2004. "We've had about 1,000 people go through by now," says William Wood, a biologist at the University of Colorado Boulder, and co-director of the programme for its first ten years.

One of the big lessons, says Wood, is that teachers should develop their lesson plans in the same way as they design experiments. Instead of following a textbook or syllabus, they should start with a clear goal — the concepts and skills that they want the students to learn. Then they should choose the instructional methods that will achieve that goal, as well as the methods they are going to use to assess the students' progress.

The summer institutes have undoubtedly done a lot to raise people's awareness, says Leupen, but simply attending a workshop is not enough. "If they come back to the same department full of the same people doing the same things — they will go right back to teaching the old way."

Continuing the support back home is crucial, says Lue. "Science is a team sport. If you're a neuroscientist or a soft-matter physicist, you have seminars, colloquia, luncheons — places you can go on campus to meet with like-minded people and trade best practices." His centre at Harvard is trying to create that kind of community for teaching and learning, he says, and many other US universities are doing the same.

As helpful as such efforts are, they do not get at what many regard as the biggest challenges for active learning. One is the lack of coordination across borders. Educational innovations are clearly happening across the globe (see page 276) and interest in active learning is high. Dirks has received an enthusiastic response to workshops she ran on pedagogy and the responsible conduct of science in Jordan, Turkey, Egypt, India and Malaysia. And leaders such as Wieman say that they regularly get invitations to talk in Europe, Asia and Australia.

Even so, the international flow of ideas is only a fraction of what it could be. The education systems differ between nations, and it is not obvious how lessons learned in one place can be applied elsewhere. What is more, researchers who are trying to get innovations into universities tend to publish in their own languages.

And then there is what Wieman and others regard as the most fundamental obstacle: the university incentive system. Too often, they say, publications and funding are the only things measured for promotion and tenure decisions, which in effect penalizes time spent on classroom innovation. "Until we commit to having teaching be a key role in tenure decisions," says Lue, "we're just paying lip service."

Some scientists say that the increased intellectual respectability of good teaching is beginning to make itself felt. Many small universities such as the UMBC are making it a key part of hiring and promotion decisions. "In our department, you don't get teaching points in tenure decisions unless you've been innovating," says Leupen. Even large research-focused universities such as Harvard are beginning to place more emphasis on instruction. "Because there is a movement, and programmes to impart this knowledge," says Dirks, "people are starting to get it." ■

M. Mitchell Waldrop is a features editor for *Nature* in Washington DC.

- Freeman, S. et al. *Proc. Natl Acad. Sci. USA* **111**, 8410–8415 (2014).
- President's Council of Advisors on Science and Technology. *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics* (White House, 2012).
- Hake, R. R. *Am. J. Phys.* **66**, 64–74 (1998).
- Singer, S. R., Nielsen, N. R. & Schweingruber, H. A. (eds). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (National Academies Press, 2012).
- Kober, N. *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering* (National Academies Press, 2015).
- Andrews, T. M., Leonard, M. J., Colgrove, C. A. & Kalinowski, S. T. *CBE Life Sci. Educ.* **10**, 394–405 (2011).



CHRISTOPH WEHRE/STIFTUNG HAUS DER KLEINEN FORSCHER

INNER SCIENTISTS UNLEASHED

Educators worldwide are experimenting with new ways to teach future researchers — from preschool onwards.

The five-year-olds are confident: trees, they agree, make the wind by shaking their branches. Their teacher does not correct them, but instead asks whether anyone has seen the wind in a place where there are no trees. One boy recalls a visit to the seashore, where the wind was whipping up water and sand with no trees in sight. Another child says that moving cars make fallen leaves twirl. Perhaps, they decide, trees are not the source of a breeze.

So goes a typical day for participants of Germany's *Haus der kleinen Forscher* (Little Scientists' House), a programme that in less than a decade has grown to reach about half of that country's children between ages three and six. Launched in 2006 by a group of German business leaders who were dismayed by their country's lacklustre performance on international student exams, Little Scientists' House got support and funding from the federal government in 2008. Today, versions of the programme are also operating in Australia, Austria, the Netherlands,

BY MONYA BAKER

Brazil and Thailand — including more than 14,000 centres in Thailand alone.

Little Scientists' House is just one of many programmes around the world that try to inspire young people's inner scientists through active engagement with the world around them. The effectiveness of this approach has been verified by hundreds of empirical studies. "It means learning content not as something you memorize and regurgitate, but as raw material for making connections, drawing inferences, creating new information — learning how to learn," says Jay Labov, a senior education adviser at the US National Academy of Sciences, one of many organizations to endorse this mode of learning. Here, *Nature* profiles innovative exemplars of such engagement, from preschool to university. If someone wanted to turn a toddler into a scientist for the twenty-first century, this is what the curriculum might look like.

Children at a German kindergarten have just found out how to make their 'bottled tornado' work.

PRESCHOOL EXPERIMENTERS

Little Scientists' House marks a departure from educators' traditional role, says Christina Jeuthe, a kindergarten teacher who participates in the programme. "You have to be willing to do something with the kids that might not lead to a result," she says. "They will not take something home that they can show their parents." Instead, teachers trained in the method try to get children to ask questions about natural phenomena and everyday objects. And when the children give naive answers (for example, that shaking leaves produce wind), the teachers help them to come up with activities to test those answers — in effect, emulating how grown-up researchers do science. But just as with scientific discovery, the end points are uncertain, says Jeuthe. "I myself had to be strong enough to not put my expectations on a specific scientific question for the kids — but let them decide, ask and discover."

In a unit about water, for example, one five-year-old argued that more water drops could collect on a euro coin than on a slightly larger 50-cent piece because the former buys more. He and his classmates counted how many drops they could dribble onto the coins' surfaces. In the end, the children could not come to a definitive answer, but that is OK, says Jeuthe. The point is to spark questions, and a conviction that they can be explored rationally.

Activities start with objects and experiences that children are familiar with — which can call for considerable creativity when adapting the programme to different places and cultures. The Australian version cannot draw on children's experience of wintry weather; instead, they focus on ice cubes. In Thailand, one activity relies on sky lanterns — miniature hot-air balloons that are common in holiday festivities. However it is done, the children say that they have fun carrying out their impromptu experiments — and in the process, say advocates of the programme, the children are learning invaluable lessons on how to plan and solve problems, not to mention gaining self-confidence.

Unfortunately, pinning down the programme's effects on students will be hard, warns Mirjam Steffensky, a chemistry educator at the Leibniz Institute for Science and Mathematics Education in Kiel, Germany. If nothing else, she says, comparisons are difficult because educators in each location are free to implement the Little Scientists' House curriculum in different ways. Still, the German Academy of Science and Engineering and other education foundations have commissioned Steffensky and several other researchers to carry out independent assessments of the programme. The three-year studies, which include control groups, will cover hundreds of students from dozens of centres to see whether the programme boosts children's language and science skills.

These assessments will not be completed until next year, but a 2013 questionnaire of more than 3,000 participating educators found that they felt more confidence and interest teaching science. "Just give the children the room, the time and the possibility," says Jeuthe. "Believe that they will work it out, and they will."

HIGH-SCHOOL COLLABORATORS

The Hwa Chong Institute (HCI) is an elite high school in Singapore that enrolls only the best-performing students and then gives them access to advanced equipment, including an atomic force microscope and cell-culture incubators. The tools would be the envy of many a university, but to director of studies Har Hui Peng, that is not enough. She has always wanted to give her students an extra challenge, and a flavour of doing science in an interconnected world. She got her chance a decade ago thanks to a lucky encounter with George Wolfe, a US educator who told her that he was setting up the Academy of Sciences (AoS): a selective, publicly funded high school in Sterling, Virginia, where students could design and conduct research. Both recognized a unique opportunity to teach their students a skill essential for twenty-first-century science: collaboration.

Every October or November since 2006, a dozen or so 14- and 15-year-old HCI students have travelled to the AoS to start research projects that will last the academic year. They work

in teams of four — two students from each country — on projects such as screening maggots for antimicrobial compounds. Nine months later, the AoS students join their HCI teammates back in Singapore to complete the final analysis and prepare presentations of the results.

Particularly at the beginning, some of the cultural stereotypes applied, says Ashley Ferguson, who took part in the programme as an AoS student. The US students were "more creative and free-flowing", she says, whereas their HCI teammates were more focused and directed: they considered what instruments were available and what experiments

could be designed around them. "Some of that more-structured thinking was good for us to learn," says Ferguson, now a senior student at the University of Virginia in Charlottesville.

Ernest Chen, an HCI graduate now studying at the University of Cambridge, UK, says that the project taught him the importance of communication. When he hit a snag with his project — chemically modifying a polymer to sop up dissolved metal ions — he and the other HCI student in his team wanted to change the methods. This

annoyed their AoS teammates, who wanted to stick with the agreed protocol. The resulting e-mail exchanges taught everyone the skills of persistence and persuasion. "Instead of just sending a first e-mail saying, 'I'm going to change this', I would say, 'we tried this, and it doesn't work, therefore we want to change it.'" Several years later, the team still stays in touch over social media.

Most important is learning to work effectively as a team, Har and Wolfe agree. The best part is when the students "start to care for each other", says Har. For example, students at one school will make sure their part of a project is completed well before another schools' exams to give their colleagues time to study, she says.

Such consideration is exactly the point, says Wolfe, now director of the AoS. "Our mission is to teach kids to do science. If you look at what scientists really do in the real world, people don't work in a vacuum."

TEENAGE RESEARCHERS

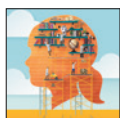
Cal Hewitt does his physics calculations by accessing a grid of distributed computers set up in the United Kingdom by CERN, the European particle-physics lab near Geneva, Switzerland. Tapping into the equivalent of nearly 40,000 personal computers, Hewitt and his colleagues are calculating the types, energies and trajectories of particles detected by an experiment developed at his institution and launched into space last year. The group's findings could suggest ways to prevent damage to satellites, and perhaps firm up theories about the source of extragalactic cosmic rays. And with any luck, this will happen before Hewitt turns 18.

Hewitt is a student at the Simon Langton school in Canterbury, UK, where students routinely design and perform real, ambitious experiments. Some of the students — Hewitt included — have presented their work at scientific conferences; a few have even published original research in the peer-reviewed literature.

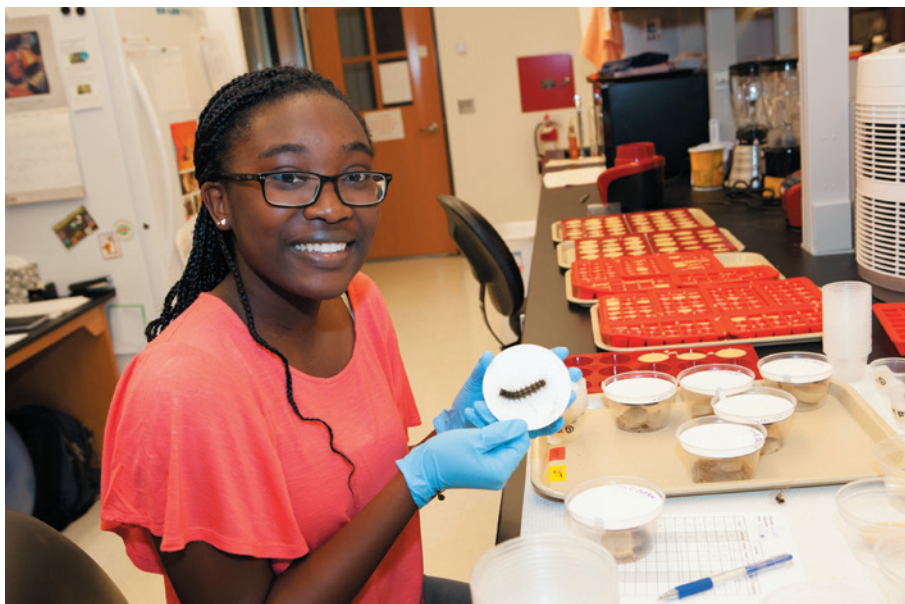
The school's philosophy is simple, says Becky Parker, who directs the Langton Star Centre, which hosts the school's research programmes: "Let's give students a chance to do real science and get the thrill of discovery."

Simon Langton is a state-funded, elite institution: students are accepted on the basis of an aptitude test at the age of 11. But the school's path to teen research began just over a decade ago, when Parker decided to sign up for a programme that gave secondary students remote access to telescopes in Australia and Hawaii. Rather than opting for the standard teacher-led

"I MYSELF HAD TO BE STRONG ENOUGH TO NOT PUT MY EXPECTATIONS ON A SPECIFIC SCIENTIFIC QUESTION FOR THE KIDS — BUT LET THEM DECIDE, ASK AND DISCOVER."



THE 21ST CENTURY SCIENTIST
A Nature and Scientific American
special issue nature.com/stem



Summer lab research for a student on the University of Richmond integrated science course.

demonstration, Parker handed the reins to her students — who used their freedom to confirm the presence of half-a-dozen known asteroids with orbits that bring them near Earth, and went on to discover two new ones.

Around the same time, Langton students entered a competition run by the UK National Space Centre to design an experiment that would be conducted in space, basing their proposal on cosmic-ray-detection technology that they had encountered on a field trip to CERN. Contest organizers offered to launch the programme if students found the funding for it. They did. And high-calibre research projects have topped students' extracurricular activities ever since.

Now the students are running calculations on data from CERN's MoEDAL (Monopole and Exotics Detector at the Large Hadron Collider) to look for some of physics' most exotic phenomena, such as microscopic black holes. Langton is the only secondary school participating as a full member in any major particle-physics collaboration, says James Pinfold, a particle physicist at the University of Alberta in Edmonton, Canada, and spokesperson for the MoEDAL collaboration. "This work in space convinced us they could handle the job," he says.

Elsewhere in the school, one student team is using genetic analysis to breed and evaluate drought-resistant strains of wheat. Another is unravelling molecular mechanisms for multiple sclerosis — a project that required a licence for genetic modification of yeast so that the students could investigate the human gene for myelin basic protein. Langton is the first secondary school to get such a licence.

Parker estimates that Langton supplies almost 1% of all students, and at least 2% of female students, who enter undergraduate physics programmes in the United Kingdom.

Other secondary schools also promote student-led research. But the scale, scope and quality of the work at the Langton centre make it stand out. To support the work, Parker and her students have raised funds from bodies such as local government and national science organizations. Such awards even supported a particle physicist to work on Langton's campus full time to advise students and build research capacity at other secondary schools.

Most credit Parker for the school's scientific success. (At one point, project teams were limited to the number of students she could fit in her car.) But Parker says that teachers are eager to put in the time for extracurricular research once they see what is possible.

To help the Langton idea spread, Parker's next project will be the Institute for Research in Schools, which will support school science teachers who want to launch genuine research projects.

And that is what education should be, says Caitlin Cooke, a Langton student who works on the MoEDAL team. "Because we've already experienced so much work at the frontier, it demonstrates to us the reality of what it is to do physics." Her colleague, Fleur Pomeroy, agrees. "Why do people question why we can be doing real science?"

INTERDISCIPLINARY UNDERGRADUATES

When Tyler Heist was considering his first year at university, he decided to throw himself into science with abandon. Most university science courses are run by individual departments and focus on a single discipline. But the Integrated Quantitative Science class at the University of Richmond in Virginia offered simultaneous introductions to five: biology, chemistry, physics, mathematics and computer science. Better still, the course would organize the lessons around interdisciplinary problems such as antibiotic resistance and cells' responses to heat. In 2010, Heist applied for one of the course's 20 available spots and was accepted. Inspired by

that experience, he will head off later this year to do doctoral work in computational biology at Princeton University in New Jersey.

The origins of the integrated course stem from a report issued more than a decade ago. The US National Research Council concluded that biological research had changed dramatically to incorporate physical and computational sciences, but biological education had not. April Hill, a biology professor at the University of Richmond, thought that the best way to fix that problem was to retool the introductory courses to view core concepts from many disciplines through the lens of real science questions, rather than taking students on the traditional march through the disciplines one by one. Hill and her colleagues ran their course for the first time in 2009.

Although interdisciplinary courses are hardly new, Hill's approach stands out for combining five distinct disciplines, for targeting introductory classes, and for including a stint of paid laboratory research in the summer following the course. Ellen Goldey, who chairs the biology department at Wofford College in Spartanburg, South Carolina, says that the University of Richmond effort has inspired other undergraduate institutions to set up similar programmes. "There is an existing model now so they will not need to reinvent the whole wheel," she says.

Hill says that the extra effort required to integrate multiple disciplines more than pays for itself; the course has prompted cross-disciplinary collaborations in her own work, on gene networks that govern the development of the most basic multicellular creatures. "Now that I have six years of interdisciplinary teaching I can't imagine not doing it," says Hill.

In 2012, the number of students taking interdisciplinary courses doubled at the university, as did efforts to recruit students from a minority background. A companion programme called SMART, now in its second year, serves students with less rigorous high-school preparation. A pre-college summer programme full of mentoring and maths helps to prepare students for the interdisciplinary courses. More than 30% of the students who took the integrated class in 2009 and 2010 went on to PhD programmes. Those who take the integrated course are more likely to graduate with a STEM major — 92% versus 60% or less of other undergraduates who start out in STEM. And they also take a greater variety of classes.

Heist, for example, says that the programme helped him to get through upper-level classes that required him to read primary biology literature that incorporated concepts from physics or computer science, and credits the course with broadening his approach to scientific investigation. "It makes you rethink the boundaries you put on things," he says. ■

Monya Baker writes and edits for *Nature* in San Francisco, California.

UNIV. RICHMOND

COMMENT

EDUCATION Improve university teaching with funds, support and data **p.282**

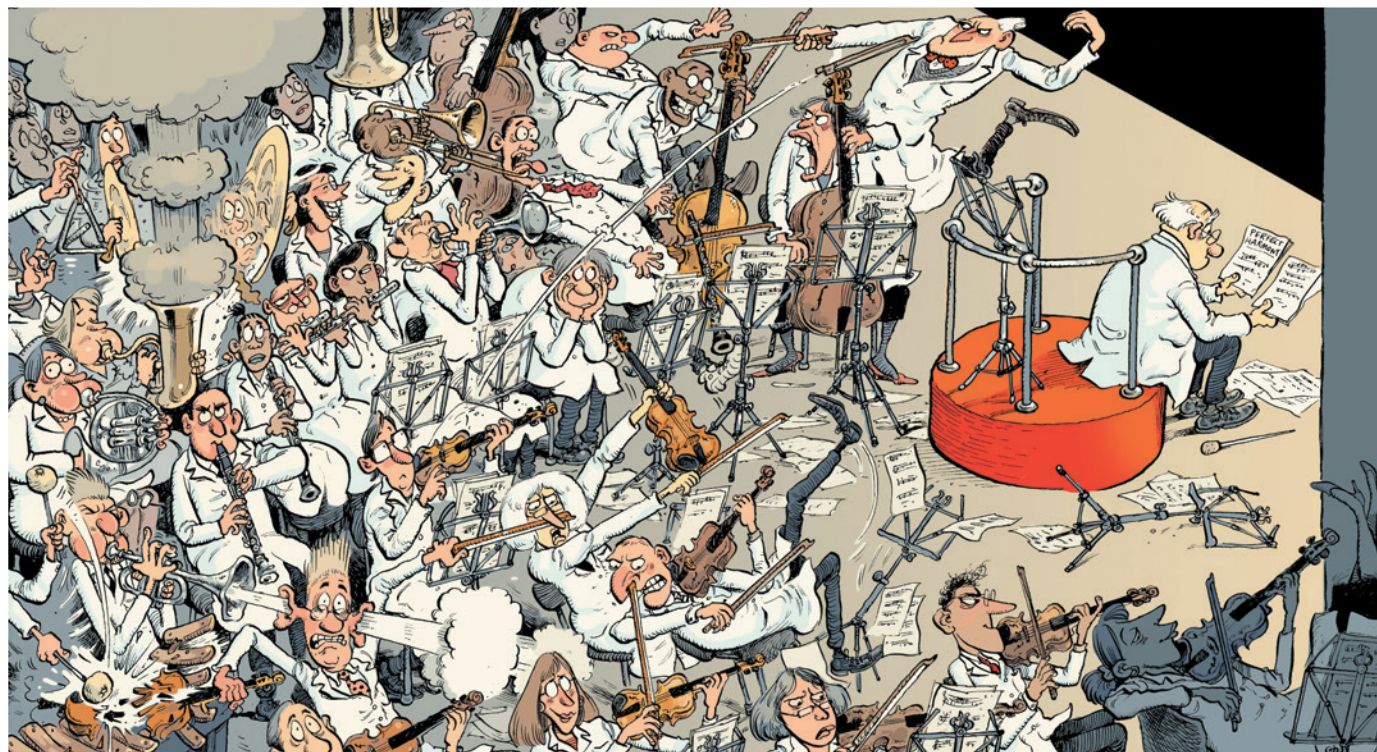


LEARNING Enable young scientists to get to grips with the wild **p.286**

CONSERVATION Clamp down on illegal trade in tiger parts could be harming lions **p.290**

HEALTH Call for a crack down on unaffordable essential medicines **p.290**

ILLUSTRATION BY DAVID PARKINS



Science professors need leadership training

To drive discovery, scientists heading up research teams large and small need to learn how people operate, argue **Charles E. Leiserson** and **Chuck McVinney**.

Education does not stop. Professors must update and develop their technical skills throughout their careers. But as they progress, few take the time — or are offered the opportunity — to become educated in how to be an effective leader.

As a consequence, academic teams waste time dealing with unproductive interpersonal issues, lack of motivation and unnecessary conflict. When things do not run smoothly, the costs in terms of money, productivity and retention of talent are high¹.

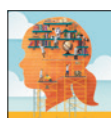
Leaders should inspire others to achieve clearly articulated, shared goals. Professors

head research teams and manage teaching staff. They lead intellectually, charting directions for advances in engineering and science that benefit society.

And the importance of these leadership skills grows as scientists gain in seniority. Even well-meaning senior professors can wreak havoc by throwing their power around and failing to take into account the

emotions of others or their own. Equally, principal investigators taking too much of a back seat can result in teams being less than the sum of their parts.

Take this true (sanitized) scenario. A major university laboratory wanted to replace their retiring director. There was no doubt as to the successor — the energetic and popular assistant director was a shoo-in. At the first meeting of the search committee, made up of a few senior lab members, the chair reviewed the procedures, which included soliciting opinions from the rest of the faculty. A consensus quickly emerged ►



THE 21ST CENTURY SCIENTIST
A *Nature* and *Scientific American*
special issue nature.com/stem

► that this ‘bureaucratic process’ would be a waste of time. “We know what the answer will be,” they said. “Everybody likes him. Let’s just appoint him now.”

See the committee’s blind spot? They were threatening to marginalize the rest of the lab, particularly junior faculty members, by failing to get their buy-in for the appointment. Instead of saving time, this high-handed behaviour could have degraded the collegiality of the lab and required needless effort to deal with the fallout. A professor who feels disenfranchised is less motivated to help solve lab issues, leaving more work for others. If they depart for greener pastures, the rest of the faculty must hire a replacement, cover the lost professor’s classes and take responsibility for abandoned graduate students. When emotions are involved, what seems like expediency can turn out to be the opposite.

In this case, one member of the committee did show true leadership, even though she had no official leadership position. She explained the risks of the rash action and persuaded a majority of the committee that the ‘bureaucratic process’ was a necessary step. The faculty interviews identified major issues for the next lab director to face, and when the popular assistant director was promoted as expected, he had a mandate for instituting important changes.

LEADERSHIP LESSONS

Over the past dozen years, we have taught leadership workshops for hundreds of engineering and science faculty members. Hardly any of the professors had ever taken a class in leadership skills or knew of any other programme similar to ours. Those who had had leadership education learned it in industry. US corporations spend about US\$14 billion each year on educating their employees in leadership and management (see go.nature.com/2kgaya). But whereas universities welcome business people taking management training courses, leadership — a word synonymous with administration and manipulation — seems to be a dirty word when it comes to their own faculty members.

Being a professor is a human-centred activity. We work with people. We teach students in classrooms, mentor our PhD students, collaborate with peers and try to persuade people in funding agencies to give us money. But leading people can be difficult, because people are not entirely rational². At most universities, junior faculty members must learn leadership skills on the job by trial and error, to the detriment of their students and careers. Senior faculty members may not understand that a failure to provide a supportive and collegial culture harms the reputation of their department or laboratory, and that they may be ill-equipped to engage effectively in large collaborative projects, such as those that dominate genomics and particle physics.

We call on academic institutions to invest in developing their professors’ human-centred leadership skills.

BACK TO SCHOOL

We met in 1999. One of us (C.E.L.) had taken a two-year leave from Massachusetts Institute of Technology (MIT) in Cambridge during the Internet boom to serve as director of system architecture at the MIT start-up Akamai Technologies. Most of the firm’s original 100 engineering staff were recruited directly from MIT and other top universities.

At the start, these brilliant academics were totally dysfunctional as a team. Every interpersonal issue you can imagine

“Research teams are best formed from a mix of diverse thinkers.”

arose: alienation, anger, apathy, arrogance, belligerence, contempt, despair, disgust, disrespect, envy, exasperation, fear, hate, impatience, indifference, jealousy, outrage, resentment, self-righteousness, spite, suspicion, vindictiveness — the whole gamut. Despite their intellectual prowess, these erstwhile academic colleagues could find no way out of this emotional morass. Many worried that they had made the wrong move in leaving academia. Morale was low.

Fortunately, Akamai’s vice-president of human resources, Steve Heinrich, supplied the right medicine. He brought in the other of us (C.M.), an experienced management consultant, to run an intensive leadership workshop for the technical leaders. Topics included dealing with emotions in the workplace; working effectively with people who think differently from you; fostering creativity; resolving conflicts; giving effective feedback; learning to recognize when different situations call for different leadership strategies; and understanding how learning curves relate to motivation. The results were immediate: harsh feelings dissipated, the engineering staff began to cooperate and technical successes started to pile up.

Back at MIT, we wondered why these ‘soft’ leadership skills were not being taught to engineering and science professors. The same kinds of emotional issues arise in university labs as in corporate workplaces. Although professors pride themselves on their rationality, they have feelings, too.

So, the two of us teamed up to adapt materials normally used for corporate training to the academic context. We also developed university-specific content from scratch, including role-playing activities involving professors and funding agencies, professors and peers, and professors and students.

We offered the workshop for the first time in 2002 to a computer-science lab (C.E.L.’s) at MIT. The response was so positive that we expanded participation to include the

Electrical Engineering and Computer Science department, and eventually, the School of Engineering and the School of Science. In 2007, we offered our two-day workshop to professors outside MIT (see shortprograms.mit.edu/lsf).

Hundreds of professors in the United States and several other countries have now taken our workshop at MIT and through custom offerings at the University of California, Berkeley; Purdue University in West Lafayette, Indiana; Harvard University in Cambridge, Massachusetts; and the National University of Singapore. Participants often express amazement at what a little leadership education can do, from reducing the number of hours spent on interpersonal issues to supplying tools for motivating students.

Our workshop focuses on how people can work together effectively. It promotes self-awareness of personal styles of leadership and offers participants new approaches to explore. Through interactive activities, self-assessment exercises and group discussions, attendees develop a repertoire of strategies for addressing common situations such as how to pitch your research programme to people outside your discipline.

Because leadership styles are individual and situational, we are careful not to judge styles as good or bad, focusing instead on helping participants to see that there may be more options available than they realized. For example, although graduate students sometimes respond well to in-depth coaching from their adviser, there are times when over-involvement can be suffocating, such as when students are starting out and need some space to get their bearings.

Participants practice their skills. For example, the module on conflict resolution concerns a dispute between two students on first authorship. One participant plays the part of the professor trying to resolve the dispute. Method-acting techniques encourage the participants playing the students to empathize with their characters, making the activity as close to a model of a real-world situation as it can be in the classroom, emotions included.

THINKING DIFFERENTLY

We use the Herrmann Brain Dominance Instrument (HBDI)³, a self-assessment survey, to explore participants’ mental diversity. Most people think of diversity in terms of the first three things that psychologists say people notice when meeting someone new: race, gender and age. But there is probably more diversity in how people think than in any physical aspect of their being.

Creativity researcher Ned Herrmann originally developed the HBDI in 1979 when he was leading management education at the General Electric conglomerate.

Herrmann was inspired by neuropsychologist Roger Sperry's work on 'split-brain' patients⁴, which showed that different areas of the brain perform specific functions (Sperry shared the 1981 Nobel Prize in Physiology and Medicine for the work). In most people, the left hemisphere is associated with speech and symbol manipulation, whereas the right hemisphere processes images and responds to sensory experiences and non-verbal clues.

Herrmann augmented Sperry's left- and right-brain metaphor to incorporate the part that emotions play in thinking. Emotions sway intellect, and intellect tempers emotions⁵. The resulting 'whole brain' model categorizes thinking styles in four quadrants (see go.nature.com/jfbqky). Left-brain thinking includes rational and safekeeping processes; right-brain thinking includes feeling and creative processes. Of course, human thought is much messier, but this approximation is helpful for understanding communication and conflicts among people.

For example, a professor can use such knowledge to 'up the game' of her research group. She realizes the advantages of matching a student's role in a project to his thinking preferences rather than to her own. Suppose that a laboratory experiment requires detailed accounting and focused individual work. A student with strong safekeeping preferences is likely to be happier and more productive in this role than a student whose preferences incline them towards interpersonal relationships. When matched to their thinking preferences, students are more likely to be motivated, to work happily and efficiently, and to self-manage, leaving more time for the professor to focus on her other priorities.

TEAM SCIENCE

Research teams are best formed from a mix of diverse thinkers. Most real-world tasks require contributions from all four quadrants. When too many people on a team exhibit the same preference patterns, they tend to compete for the same 'desirable' roles, and it can be hard to find someone to do the 'undesirable' chores. A diverse team gives everyone a chance to contribute in a complementary fashion. And research shows⁶ that gender-balanced teams of diverse thinkers tend to outperform same-thinking teams.

Professors tend to be sceptical about many things, and leadership is no exception. Over the years, we have heard many academic colleagues in engineering and science, especially senior ones, express opinions as to why soft skills are pseudoscience and should not be taken seriously: "people skills cannot be measured and understood the way that a subatomic particle, a strand of DNA or a computer algorithm can be"; "humans are unpredictable and emotional and cannot be understood systematically"; and "people

skills are unimportant in the academic world because everyone tends to act rationally". It is no wonder that so few universities have bothered to teach leadership skills to their faculty.

Although persuading professors to change is notoriously hard⁷, there are indications that things are improving. Team science⁸ is a rapidly growing cross-disciplinary field of study that aims to maximize the efficiency and effectiveness of team-based research in the sciences. The growing interest in entrepreneurship among technical academics has led to a greater understanding in universities of the importance of leadership skills. And 'big science' endeavours highlight the importance of getting many people to work together effectively. Examples include CERN (Europe's particle-physics lab near Geneva, Switzerland), ENCODE and the many '-ome' projects (such as the Human Genome Project).

But leadership training alone is not enough. Academia must support and reward leadership, embracing the modern understanding that thinking — the cornerstone of academic accomplishment — involves emotion. Engineering and science must adapt to value the quality of interpersonal relationships, which are essential to teamwork. They must respect diversity of thought, especially non-technical modes, if they wish to inspire creativity.

Smooth-functioning and innovative research teams are essential for producing the inventions and discoveries needed to address the many challenging problems that our society faces. ■

Charles E. Leiserson is professor of computer science and engineering at the MIT Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. **Chuck McVinney** is president of McVinney & Co., a management-consulting firm specializing in facilitating creative thinking and leadership, in Brookline, Massachusetts, USA. e-mails: cel@mit.edu; chuckmcvinney@gmail.com

1. Stephan, P. E. & Ehrenberg, R. G. *Science and the University* (Univ. Wisconsin Press, 2007).
2. Kahneman, D. *Thinking, Fast and Slow* (Farrar, Straus and Giroux, 2011).
3. Herrmann, N. & Herrmann-Nehdi, A. *The Whole Brain Business Book* 2nd edn (McGraw Hill, 2015).
4. Gazzaniga, M. S. *Tales from Both Sides of the Brain* (HarperCollins, 2015).
5. Goleman, D. *Emotional Intelligence* (Bantam Dell, 1995).
6. DeRidder, C. G. & Wilcox, M. A. *How to Improve Productivity: Whole Brain Teams Set New Benchmarks* (Herrmann International, 2006).
7. Bruns, J. W. & Bruns, D. L. *J. Leadersh. Stud.* **1**, 53–63 (2007).
8. Bennett, L. M., Gadlin, H. & Levine-Finley, S. *Collaboration & Team Science: A Field Guide* (National Institutes of Health, 2010).

The authors declare competing financial interests: see go.nature.com/nlpne9.

Improve undergraduate science education

It is time to use evidence-based teaching practices at all levels by providing incentives and effective evaluations, urge **Stephen E. Bradforth**, **Emily R. Miller** and colleagues.

Undergraduate students are not being taught science, technology, engineering and mathematics (STEM) subjects as well as they need to be¹. Too often, faculty members talk at students rather than engaging them in activities that help them to learn and apply core scientific concepts and skills. Despite growing scholarship about effective teaching methods and meaningful ways to assess them³, research universities rarely provide adequate incentives, support or rewards for the time that faculty members spend on improving teaching. And faculty members assign a low priority to undergraduate teaching compared to research⁴. Efforts to improve undergraduate STEM education have been slow and piecemeal at best.

The time is now ripe for change. Today we can collect, analyse and assess individual and institutional data on teaching effectiveness and student outcomes in ways not previously possible. There are also successful models for supporting and rewarding scientists to be both excellent teachers and researchers.

We write as representatives of the Association of American Universities (AAU) and the Research Corporation for Science Advancement Cottrell Scholars, a group of research-active science faculty members. We call for immediate change at all levels of research universities to improve the quality of university STEM education. It is no longer acceptable to blame primary- and secondary-school teachers for the deficits in STEM learning at the university level.

To facilitate change, we outline here rigorous examples of best-in-class pedagogical practices, programmes and policies. Many of these ideas are not new — but the robustness of implementing and evaluating them is.

Although our experiences are in the United States, these principles are more broadly applicable. No single tool will work for all universities, but every university now has at its disposal the tools to improve undergraduate STEM teaching, and no defensible reason for not using them.

Valuing teaching must move from rhetoric



MICHAEL MARSLAND/YALE UNIV.

The Yale Center for Engineering Innovation and Design is used by engineering as well as art students.

to reality. Too many students at North American universities who intend to major in STEM fields do not end up doing so, often because of the traditional teaching practices used⁵.

Active learning interventions improve achievement for all students; those with disadvantaged and ethnic-minority backgrounds gain the most⁶. Administrators and faculty members have a responsibility to ensure that introductory classes do not push students away from STEM courses but promote critical thinking, problem solving, engaged learning and knowledge retention for all students, whether they intend to major in STEM fields or not. Effective STEM teaching is crucial to developing a science-literate population that can address the complex and interdisciplinary health, energy, security and environmental challenges of our time.

CLASS ACTION

For decades, North American universities have relied almost exclusively on end-of-term student surveys of little use in assessing teaching performance⁷. Unfortunately, other nations have followed suit. Faculty members

tend to be assessed and promoted mainly on the basis of research success, which, unlike teaching, is readily quantified — through grant funding, the number and perceived importance of publications, and citation metrics. These systemic obstacles have proved resistant to change.

There are indications, however, that we are approaching a tipping point. Scientific knowledge about effective teaching methods has increased, as outlined in a 2015 report by the US National Research Council⁷, which follows up on a 2012 report³ that synthesized literature from several fields on how students learn, particularly in scientific disciplines, and ways to improve instruction.

Policy-makers are increasingly questioning the value of an undergraduate education from a large research university given its rising costs, and they are calling for accountability and efficiency measures. And faculty members are re-evaluating their teaching methods in response to competition from the increasing number and quality of massive open online courses (MOOCs) and other online offerings.



THE 21ST CENTURY SCIENTIST
A *Nature* and *Scientific American*
special issue nature.com/stem

SOURCE: UCD

Past efforts to improve teaching have focused on individual faculty members. The AAU Undergraduate STEM Education Initiative, launched in 2011, has been exploring a more systemic view of educational reform. It is based on understanding the wider setting in which educational innovations take place — the department, the college, the university and the national level. Thus, it emphasizes the separate roles of senior university administrators (who can implement top-down change), individual faculty members (bottom-up change) and departments (change from the middle out), all of which are necessary for sustained institutional improvement to undergraduate STEM teaching and learning⁸.

We suggest strategies that target changes at each of these levels, inspired by what we have learned from the initiative and from individual faculty members' experiences.

BOTTOM UP: FACULTY MEMBERS

Effective teaching begins with faculty members, who maintain significant autonomy over their practices. Most care deeply about teaching, in addition to their strong interest in research. However, they rarely obtain informative feedback about their students' learning and often are unfamiliar with improved teaching practices. Moreover, they lack the resources and support to devote a significant proportion of their time to change their current practices.

Increase scientific and reflective teaching.

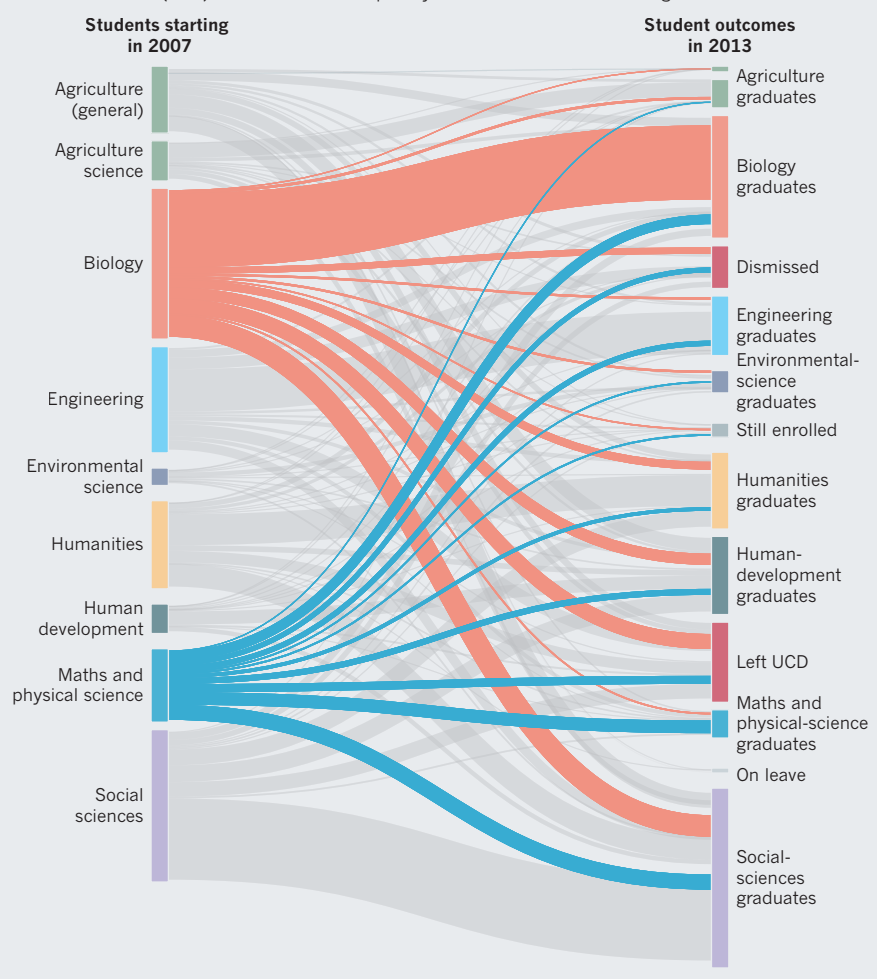
Faculty members need to shift their perspective from "What did I teach?" to "What did my students learn?" They must aim to create engaging learning environments in which students are participants rather than passive note-takers or followers of 'cookbook' laboratory experiments. Faculty members should experiment with evaluation strategies such as classroom observation protocols and pre- and post-course testing, which can guide mid-course adjustments. The Yale Center for Scientific Teaching in New Haven, Connecticut, has demonstrated the value of such strategies, and trains faculty members, instructors, postdocs and graduate students in their use.

Increase student engagement in learning.

Students learn better when they participate in and reflect on their own learning process. One way to assess students' project ownership is linguistic analysis. In responses to open-ended course-evaluation questions, the use of first-person personal pronouns and emotional words — for example, "my research made me excited about science" — correlate with student ownership of learning. Excellent teachers facilitate this, and it is one of the psychosocial factors involved in retaining students in the sciences⁹. Exemplary

CHARTING A PATH

Visualizing students' educational journeys has informed recruitment and retention efforts at the University of California Davis (UCD). The tool was developed by the iAMSTEM HUB in Undergraduate Education.



programmes using this tool to measure student engagement in STEM education are the University of Texas at Austin's Freshman Research Initiative and Yale's Rainforest Expedition and Laboratory courses.

TOP DOWN: SENIOR ADMINISTRATION

Beyond the classroom, institutional infrastructure — physical and organizational — is required for effective teaching. Senior university administrators have crucial roles in creating a culture that values teaching and a support structure that drives continuing improvement and innovation. Value and mission statements are not enough.

Recognize and reward good teaching. To influence how faculty prioritize their efforts, university administrators must promote excellent teaching at all levels. A powerful example of this concept is the establishment of endowed chairs for educational excellence in STEM departments, such as the professor of STEM education in the chemistry department at Washington University in St. Louis, Missouri. This chair acknowledges and funds

faculty members' innovative research and teaching. Industrial and alumni benefactors can support such chairs.

Encourage faculty buy-in. Senior university administrators must support deans and department heads with professional development, assessment tools and other resources to improve teaching. This is crucial if faculty members are to step out of their comfort zones and introduce proven teaching techniques, some of which students may initially resist, being accustomed to traditional 'chalk-and-talk' lectures. The Nucleus programme at the University of Virginia in Charlottesville seeks proposals from departmental teams to promote significant and sustained change in introductory STEM courses. Funds are awarded to both the faculty members and chairs to support the changes.

Centralize and make accessible data and analytics. Universities accumulate volumes of longitudinal data that have been underused in assessments of student learning and degree progress. Robust, scalable and centralized

campus-wide analytics leverage existing data and reduce the need for multiple assessment tools. For example, the iAMSTEM HUB at the University of California, Davis, offers tools to visualize student pathways through courses and programmes. At its most basic level, the ribbon-flow tool (see 'Charting a path') informs recruitment and retention efforts by visualizing the starting and ending disciplines for student year groups.

Findings should be shared with departments and faculty members to inform discussion and action. The iAMSTEM HUB formed a community of institutions to share analytical and visualization tools, called Tools for Evidence-Based Action. The Committee on Institutional Cooperation is an academic consortium of 15 research-intensive universities that includes the University of Michigan in Ann Arbor and the University of Wisconsin–Madison. It has embarked on a learning and research analytics project and shares results on a common set of student success measures.

Use teaching improvement as a fund-raising lever. Alumni, as well as private and public-sector employers, have direct interests in enhancing the university teaching and learning experience. Senior administrators should incorporate into their fund-raising campaigns well-articulated initiatives to improve STEM education.

MIDDLE OUT: COLLEGES AND DEPARTMENTS

Improvements at the top and bottom are sustainable only if combined with changes at the colleges (subsets of universities) and departmental levels that foster a team culture of continuous teaching improvement.

Develop learning objectives for introductory STEM courses. Curriculum design must include appropriate learning objectives for introductory service courses and those required for major subjects. For example, the Force Concept Inventory — broadly adopted by the physics discipline — identifies core concepts that should be mastered by all first-year undergraduate physics students. Faculty members need to align their course curricula to these departmental learning goals, in both content knowledge and skill development.

For example, Michigan State University in East Lansing is changing how introductory biology, chemistry and physics are taught, by redesigning course curricula and assessments. Faculty members have debated core disciplinary ideas, cross-cutting concepts and scientific practices. Each discipline has generated a list to transform the introductory curriculum and is now developing assessments for all three dimensions of learning.

The chemistry department at the University of Arizona adopted a course called Chemical Thinking, a redesigned general-chemistry module that has improved students' learning,

retention of information, and performance in advanced courses. The course focuses on group discussions and problem solving. Faculty members spend less than ten minutes lecturing in an hour-long class.

Give faculty members the time and resources to improve teaching. Teaching assignments should accommodate the challenges and time necessary to improve methods and integrate new assessment techniques. Departments should reallocate funds to support teaching innovation and encourage staff to use campus centres for teaching and learning. At the University of Kansas in Lawrence, the STEM departments and the Center for Teaching Excellence have partnered to fund teaching postdocs to help redesign courses.

Encourage peer support and cross-departmental dialogue. Colleges and departments need to stimulate discussion of

"It will require a reallocation of funds — not just priorities."

and respect for teaching. Such discussions can be strengthened through co-teaching arrangements, in which two people share course instruction and mentor each other. This approach has been successful in the College of Biological Sciences at the University of Minnesota, Twin Cities, and the University of North Carolina at Chapel Hill. Both co-teachers receive full workload credit.

Evaluate teaching with meaningful metrics. Departments and schools should supplement alternative student evaluations with metrics for teaching performance. Methods can be chosen on the basis of the needs and resources of each institution and include pre- and post-course testing, classroom observation and reflective teaching statements or portfolios. The emphasis on student evaluations must shift from reflecting the popularity of instructors to effectively assessing learning. Macquarie University in Sydney, Australia, has established a teaching index to recognize and financially reward departments for the volume and quality of faculty members actively dedicated to teaching.

Make teaching count for promotion and tenure. Review committees must be trained to evaluate, use and weigh up data on teaching and learning, following carefully crafted guidelines. In 2007, the University of British Columbia in Vancouver, Canada, launched a major science-education initiative led by Nobel-prizewinning physicist and educator Carl Wieman with the goal of transforming undergraduate science and mathematics. From this initiative emerged a tenure track for teaching faculty, whose promotion depends on excellence and leadership in education,

with much less emphasis on research. This has generated guidelines for measuring teaching effectiveness. These measures should be examined for research faculty as well.

SCHOLARLY TEACHING

Progress is likely to be slower than many of us would prefer. Moving forward will require a commitment at all levels to share and adapt the practices highlighted, to achieve systemic and sustainable change in undergraduate STEM education. It will require a reallocation of funds — not just priorities.

As a first step, institutions, colleges and departments must expect and enable their faculty members to be scholarly about teaching. And they must assess, recognize and reward those who are. ■

Stephen E. Bradforth is chair and professor in the Department of Chemistry, University of Southern California, Los Angeles, California, USA. **Emily R. Miller** is director of the AAU Undergraduate STEM Education Initiative at the Association of American Universities, Washington DC, USA. **William R. Dichtel** is associate professor of chemistry and chemical biology, Cornell University, Ithaca, New York, USA. **Adam K. Leibovich** is professor of physics and astronomy, University of Pittsburgh, Pennsylvania, USA. **Andrew L. Feig** is professor of chemistry, Wayne State University, Detroit, Michigan, USA. **James D. Martin** is professor of chemistry, North Carolina State University, Raleigh, North Carolina, USA. **Karen S. Bjorkman** is dean and distinguished university professor in the Department of Physics and Astronomy, University of Toledo, Ohio, USA. **Zachary D. Schultz** is associate professor of chemistry and biochemistry, University of Notre Dame, Indiana, USA. **Tobin L. Smith** is vice-president for policy at the Association of American Universities, Washington DC, USA. e-mail: emily.miller@aaau.edu

- Freeman, S. et al. *Proc. Natl Acad. Sci. USA* **111**, 8410–8415 (2014).
- Singer, S. R. et al. (eds) *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (US National Research Council, 2012).
- Fairweather, J. in *The Changing Face of Academic Life: Analytical and Comparative Perspectives* (eds Enders, J. & deWeert, E.) 171–192 (Palgrave Macmillan, 2009).
- Seymour, E. & Hewitt, N. M. *Talking About Leaving: Why Undergraduates Leave the Sciences* (Westview, 1997).
- Eddy, S. L. & Hogan, K. A. *CBE Life Sci. Educ.* **13**, 453–468 (2014).
- Stark, P. B. & Freishtat, R. *ScienceOpen Res.* <http://dx.doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1> (2014).
- Kober, N. *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering* (US National Research Council, 2015).
- Corbo, J. C. et al. Preprint at <http://arxiv.org/abs/1412.3034> (2015).
- Hanauer, D. I. & Dolan, E. L. *CBE Life Sci. Educ.* **3**, 149–158 (2014).
- Benton, S. L., Duchon, D. & Pallett, W. H. *Assess. Eval. High. Educ.* **38**, 377–388 (2013).



PVS/STOCK.COM/ALAMY

Immersive, dynamic and multidimensional, nature can stimulate and engage the whole child.

EARLY CHILD DEVELOPMENT

Body of knowledge

As government education experts call for toddler literacy, and baby apps proliferate, are we losing sight of materials-based learning? Infant scientists and young explorers thrive in the open air and through free play, eager to grasp the world — literally.

Carbon-copy playgrounds. Cramped classrooms. ‘Car park’ school grounds. Across the industrialized world, these are the environments in which most young children are expected to play and learn; zoo enclosures can look more enriched. Studies are emerging that reveal poor design as a hindrance to learning in the very young, as damaging as militaristic drills. Meanwhile, frogspawn and starry skies — once the recruiting agents of science — are beyond many children’s experience. Here, two leading practitioners in developing learning environments for the young set out what needs to happen.

The nexus of young child and physical world is a supercharged space, where burning potential meets a chunk of the new. Developmental neuroscientist Alison Gopnik has described young children’s lamp-like

awareness of the world, the 360° intensity of their absorption. Babies’ brains grow at the gallop, making 700 new neural connections a second. By the age of 3, a child has 1,000 trillion synapses, up to 4 times the number in an adult brain; these are later pruned.

While this neural crescendo builds, the infant and toddler is exploring the world holistically, with hands, feet and body, as well as eyes, ears, nose and mouth. Sensorimotor, materials-based learning was the bedrock of our evolution, and it shapes the brain: in his 1973 *The Ascent of Man* (Little, Brown), mathematician and biologist Jacob Bronowski wrote: “The hand is the cutting

edge of the mind.” It is people and things, Gopnik tells us — sand, water, a bean plant, sociable family members — that satisfy the infant scientist’s experimental urge.

That young children learn how the physical world and people work so tangibly — literally getting to grips with them — militates against the idea that intensive academic settings are needed to set under-sixes on course to becoming tomorrow’s synthetic biologists and solar engineers. That, cognitive scientist Guy Claxton argues, reduces learning to listening and reading, talking and writing.

E-media are often viewed as a corrective to such reductionism. But with many babies and toddlers now habitués of the virtual world — and even some nursery classrooms boasting giant digital whiteboards — electronic gadgetry is hotly debated in the early-learning context. Whatever the final verdict,



THE 21ST CENTURY SCIENTIST
A *Nature* and *Scientific American*
special issue nature.com/stem

it is clear that virtual experience alone fails to engage the whole child: screen-time limits children to what Claxton calls “eye-brain-finger” learning. It can also intrude on ‘face time’ and the socialization and language learning that are central to early child development. (Reportedly, prominent technocrats such as Twitter founder Evan Williams strictly limit their own children’s use of electronic devices.)

Meanwhile, ‘old school’ alternatives to reductionism have a long pedigree. Modern findings on the importance of physicality and social contact for the very young echo the practice of pedagogical innovators and reformers from the early twentieth century on. One of Italy’s first female physicians, Maria Montessori (1870–1952) — observing that cognition is closely tied to movement, and learning to autonomy — emphasized play- and materials-based education. Montessori classrooms have been compared to university labs, where children direct their own work at their own pace, with ‘colleagues’ of their choosing and materials designed for multifaceted experimentation. Three-year-olds, for example, might handle and sort knobbed wooden cylinders of differing dimensions, and in the process learn the ‘pincer grip’ needed for manipulating fine instruments, as well as skills including reasoning and comparison, and the rudiments of maths.

A similar philosophy permeates other systems shown to boost early language use, and to strengthen reflection and self-monitoring. Finland focuses education for children up to 7 on play-based physical and social learning — and the country’s 15-year-olds rank first in Europe in science and reading in surveys by the Organisation for Economic Co-operation and Development’s Programme for International Student Assessment.

More radical still is forest school, rooted in the ‘outdoor nursery’ model of educational reformer Margaret McMillan (1860–1931), among others, and wending its way from Britain to the United States and Scandinavia between the 1910s and the 1990s. A ‘classroom’ might be a woodland clearing; ‘lessons’, fire-building or insect identification; methodology, supervised risk-taking and close observation.

Education *en plein air* has inspired biologists from Charles Darwin to E. O. Wilson, and the curriculum carries a range of benefits key to work at the bench. A study of UK forest schools, commissioned by the Forestry Commission and run by Forest Research and think-tank the New Economics Foundation, found improvements in children’s confidence, concentration, fine motor control and teamwork. Forest schools also offer tangible evidence of abstract phenomena such as life cycles, food chains and materials behaviour (such as why wood blackens in a fire).

Nature is unsurpassed as a teaching



Günter Beltzig’s playspaces often give opportunities for learning about materials.

environment, argues social ecologist Stephen Kellert on page 288. A meadow or seashore is an immersive experience, engaging and challenging the whole child physically, socially, cognitively and emotionally. Complex and unexpected, a drama of rain, wind, birds flying, mud squelching and leaves falling, nature sparks curiosity even as it provides an arena for free play — both gripping detail and far horizons.

But with the wild giving way to traffic-ridden tarmac, safe places for full-on play (and outdoor learning) become ever more important as refuges and testing grounds for children. ‘Quality’ play that stretches a child’s senses, intellect, sociality, physicality and motor skills both gross and fine demands spaces designed to accommodate children’s needs and behaviour, notes renowned play-space designer Günter Beltzig on this page.

As educational theorists, teachers, governments and parents tussle endlessly over literacy, numeracy and testing in the under-eights, we sometimes forget that nature nurtures. We ascended as a species through incandescent curiosity — that hallmark of scientists in every century — at play in the world. Those primal, physical realities still have everything to teach us, in one way or another. **Barbara Kiser**

GÜNTER BELTZIG Learn to play, play to learn

*Industrial and playground designer,
Hohenwart, Germany.*

An urban space equipped with a slide and swings is known as a playground. Most such spaces are designed by landscape architects. There is still no academic underpinning to playground-design theory, even though there is for graveyard design. The body of knowledge on children’s behaviour gathered by educationalists, psychologists and neuroscientists has not permeated playground design, so most such spaces fail to inspire play and the physical, emotional, social and psychological development that it engenders.

Designers are trained in creating and problem solving, and in materials and production science. But to create an environment for children, you need to behave like an anthropologist, respecting the ‘otherness’ of children. You learn by reflective observation to discover their needs, possibilities and aims (see ‘Six rules for play spaces’).

PLAY FOR KEEPS

Six rules for play spaces

- Communicate that this is a place to linger — neither a training ground nor a display area for structures attuned to an adult aesthetic.
- Allow the discovery of things that expose themselves only to the seeker, such as a nook cut into a hedge.
- Give opportunities to experiment with self-controlled, obvious and manageable risk, for example with the ability to climb to different levels.
- Allow different groups to find desirable places according to mood, interests or needs.
- Provide shelter from wind, noise and observation by others.
- Avoid superfluous bans.

► Being a child is a fluid, constantly changing process. Whatever a child is capable of doing today, they might not have achieved yesterday; whatever they will be doing tomorrow, they might not do today. This ongoing 'newness', this changing of mind, this physiological development, comes in waves — not in a constant rhythm. Some develop early, others travel on a later wave.

The continuing changing of body and mind leads to a sort of constant discourse within the capabilities that the child has. This self-experimentation to find those capabilities, harnessing unbiased, spontaneous fantasy, creativity, curiosity, is what playing really is. It is the origin of learning, not just with your brain, but in combination with intuition and feelings, with your body: learning as a whole being. It is goal-free discovery through experimentation, in a world where the child is the decision-maker, not the decided-for. For this sort of learning, a child needs time, freedom of possibilities and space. For our increasingly urbanized species, that space is the playground.

The characteristics of the community will dictate playground design to a degree. Neighbourhoods differ socio-economically and demographically as well as in topography, infrastructure and degree of urbanization. Inside the play space, the layout and nature of features will determine the children's behaviour, and so demand informed design to ensure that the children's needs and wants, individually and collectively, are met.

Components for climbing and balancing — such as ladders, ropes and hanging



RICHTER SPIELGERÄTE

Play towers and bridges offer the chance to take manageable risks, such as climbing and balancing.

bridges — foster planning, problem solving, physical coordination, body control, risk management and the chance to work on reaction time, as well as an understanding of capabilities and shortcomings. Climbing equipment should always be part of a wider schema and have an objective, such as offering a shortcut or a way to a platform. In isolation, it will lead to disputes over territory.

Towers and high-level playhouses offer vistas: the experience of seeing the world differently and from above. This can feel transformative for the child, so high structures are hugely popular and demand multiple ways of climbing up and down simultaneously, as well as extremely sturdy construction.

Shared spaces allow the growth of self-assertiveness and social intelligence, but calm spaces for daydreaming and relaxing offer the chance to reflect, think or escape. Wicker playhouses, hollows in shrubs or low-level huts for quiet play should be located away from the most popular equipment.

Sandboxes should mimic natural hollows, built below ground level to prevent loss of sand. They encourage the exploration of materials and, in combination with equipment such as excavators and wagons for digging and transporting sand, provide opportunities for physical fitness, technical experience and teamwork.

Figures should be functional rather than merely decorative. Children do not necessarily share adult aesthetics. More importantly, they are interested in simple shapes, suggestive rather than representative, which can be adapted in imaginative play: a log can be a horse or motorcycle, a litter bin a spaceship.

Inclusive play spaces for children of all abilities need particular tailoring. Children with hearing impairment are limited to their field of vision, so moving components must be highly distinctive — differently coloured, for example. Children with visual impairment need safety areas distinguished by barriers or different surfacing. Wheelchair use demands close attention to surfacing and dimensions, and the placement of some play components, such as tube phones, along paths.

No playground is perfect. When I watch children play in one that I have designed, I see them use it in ways I had not thought of. To enable children to experiment freely and curiously through play, developmental psychologists and playground designers need to find a meeting place. Our urbanizing world is essentially hostile to children, so the creation of play spaces honouring what the young themselves want and need is, in my view, central to bettering global society.

STEPHEN KELLERT

Build nature into education

Professor emeritus and senior research scholar, Yale School of Forestry and Environmental Studies, New Haven, Connecticut.

The typical child in the United States now spends 90% of the time indoors. US children aged 2–5 engage in electronic media for an

FURTHER READING

THE HUNGRY MIND SUSAN ENGEL (HARVARD UNIV. PRESS, 2015)/THE SCIENTIST IN THE CRIB ALISON GOPNIK ET AL. (HARPERCOLLINS, 1999)



Machias Elementary School in Snohomish, Washington state, is an example of biophilic design.

average of more than 30 hours per week; for 8–18-year-olds the figure is 52 hours. Most children devote just 30 minutes daily to unstructured outdoor play; a generation ago, it was more than 4 hours. Many parents fear letting their children play outdoors on their own, and see learning as a formal indoor process.

A growing body of evidence from US, European and Australian studies — reviewed in Stephen Moss's 2012 publication *Natural Childhood*, produced for the UK National Trust — suggests that this disconnection may be causing physical, emotional and intellectual deficits in children's learning and development. A young child engaged in free play in a grove of trees or under a garden bush experiences a wealth of kinetic, aural, visual and tactile stimulation. These experiences foster a wide array of adaptive responses that provoke curiosity, observation, wonder, exploration, problem-solving and creativity. A child building a dam or den gains understanding of gradients, forces, materials, the behaviour of water and wood, and the local environment.

The centrality of nature in children's learning begins with our origins as a species. For more than 99% of our evolutionary history, humans adapted in response to mainly natural forces. We became inclined to affiliate with nature, a tendency called biophilia. If it is to flourish, biophilia demands

experience and nurturance. The reliance on learning is the source of our species' remarkable inventiveness, yet it also carries the potential for us to behave in ways contrary to our long-term biological self-interests.

Beyond a vague awareness that the 'outdoors' is good for kids, we are only starting to explore the role of nature in learning and development. The scientific evidence remains limited, but findings in health care, education, work, recreation and community indicate that contact with nature remains vital to child development, and it may not be possible to find a substitute. A study of 90 schools for children aged 5 to 12 in Australia, for example, found that being outdoors improved the children's self-confidence, ability to work with others, caring, peer relationships and interaction with adults (C. Maller and M. Townsend *Int. J. Learn.* **12**, 359–372; 2006).

Immersion in the sensory and informational richness and dynamic qualities of woodlands, beaches and meadows evokes basic learning responses such as identification, differentiation, analysis and evaluation. Children distinguish big trees from little trees, house plants from garden plants, vines from ferns, ants from flies, ducks from songbirds, real creatures from imaginary ones. They develop quantitative skills by counting insects and flowers; gain materials knowledge from playing in grass and mud; intuit physics from how creek water responds to obstacles and opportunities. In recognizing hills, valleys, lakes, rivers and mountains, they learn geological form. In engaging with other life from redwood

trees to hedgehogs, they encounter an endless source of curiosity, emotional attachment and a motivation for learning. In adapting to the ever-changing, often unpredictable natural world, they learn to cope and problem-solve.

Alan Ewert's classic 1989 book *Outdoor Adventure Pursuits* (Gorsuch Scarisbrick) reviewed studies of children participating in nature programmes, and found that these children asked more questions than others, and were better at solving problems. A study of 262 children aged 3–12 in poor neighbourhoods in Chicago, Illinois, demonstrated richer creative play following exposure to nature (A. Faber Taylor *et al. Environ. Behav.* **30**, 3–27; 1998).

Unfortunately, modern society has erected increasing barriers to children's contact with nature. Children's residential, educational and recreational environments, for example, are often highly artificial and sensorily deprived. A new paradigm is needed: biophilic design. This approach to building and landscape design encourages direct and indirect contact with nature, and an experience of place evoking children's evolved affinities for the natural world.

The direct experience of nature — light, air, plants, animals, water and landscapes — can be designed, for example, through abundant use of plants indoors and outdoors; subtle manipulations of airflow, temperature and pressure; and outside views. Nature-inspired artwork, the use of materials such as wood and wool, and designs that mimic natural form can provide an indirect experience of the living world. So, too, can sheltered spaces with long views and visual connections to the outside, and areas rich in natural features, yet orderly and understandable.

Biophilic design has the potential to transform children's schools and play areas. A recent example is the new Sandy Hook Elementary School in Newtown, Connecticut, under construction on the site of the 2012 shootings where 20 children died. The original school was widely seen as a barren structure. The new school, for which I was a design consultant, contains extensive natural lighting and materials, outside views, widespread planting and courtyards, and constructed wetlands. Its biophilic design features encourage learning and, just as importantly, affirm life at this site of unimaginable loss.

The natural world is more than a decorative backdrop or a dispensable amenity. Experiencing it is an act of profound self-interest, a guard against a future as imperilling to the long-term fitness of our species as the more obvious threats of poverty and disease. ■

Correspondence

Tiger–bone trade could threaten lions

The clampdown on the illegal trade in tiger parts for traditional Chinese medicines seems to be prompting their substitution with lion products (see our report *Bones of Contention* TRAFFIC/WildCRU; in the press). Our investigations also reveal that measures to protect tigers in Asia may be driving a growing trade in other Asian felids.

Images of lions started to appear from about 1995 on the labels of Chinese medicines that traditionally would have contained tiger products. In December 2009, a CITES permit was issued in South Africa for lion skeletons to be exported to Asia — indicative of a developing market for the bones. Our report notes that more than 1,160 skeletons had been exported by the end of 2011, of which 85% were destined for the Lao People's Democratic Republic.

Controversially, the trade in lions from South Africa seems mostly to be a sustainable by-product of the trophy-hunting industry (hunted lions are almost always bred in captivity). However, anecdotal accounts of wild lion poaching in other African countries for markets in Asia call for further investigation. And the trade in lion bones from South Africa may be linked with the sharp increase in rhino poaching in 2008, as discussed in our report. **Vivienne L. Williams*** *University of the Witwatersrand, South Africa.* vivwill@netdial.co.za

**On behalf of 4 correspondents (see go.nature.com/gvklj for full list).*

Mind the gap in Hunt responses

In my view, you use your publishing might irresponsibly in presenting Alessia Errico's article 'Judge by actions, not words' as the sole female riposte to Nobel laureate Tim Hunt's misogynistic

remarks (*Nature* 522, 393; 2015).

Besides running contrary to the groundswell of reaction from many scientists (see, for example, go.nature.com/gdvzv), Errico's statement of support for Hunt's remarks was predictable. Hunt was her former supervisor: she depended on him for scholarship and career advancement.

Whether or not Hunt was joking and whether or not he apologized satisfactorily are beside the point. Neither is it likely that such outdated and seemingly entrenched attitudes can be dispelled by practical attempts to counter gender inequality in science (see *Nature* 522, 255; 2015 and D. Hilton *Nature* 523, 7; 2015).

Conspicuous by its absence in *Nature* so far is this: a woman commenting on the harm done by the flippant public denigration of women in science by a prominent scientist who is male. **Rebecca Williams Jackson** *Big Rapids, Michigan, USA.* rebecca.williams.jackson@post.harvard.edu

Make pharma justify the price of drugs

Some 500,000 people worldwide die from hepatitis C virus every year because available remedies are unaffordable. So it is good to see expensive medicines against this virus, along with anticancer agents, included on the latest World Health Organization (WHO) Model Lists of Essential Medicines (see go.nature.com/wzah9w). However, simply listing these drugs is unlikely to make them more affordable, particularly in the developing world (see also www.msfaccess.org).

Health authorities must drive home the ethical obligation of pharmaceutical companies to cut the cost of life-saving medicines in accordance with the Doha Declaration of 2001, especially given that even some wealthy countries cannot afford them. To support this, governments

and national health authorities should not grant patents on medicines that could be made more affordable.

Authorities should insist that drug firms make their costing information publicly available. This would include their investment in research and development compared with marketing expenditure, and details of differences between the actual cost and the retail price of medicines.

Such measures would enable the public to judge to what extent drug prices are the result of investment in innovation, rather than of satisfying shareholders and supporting marketing.

Vittorio Berteletti*, Silvio Garattini *IRCCS — Mario Negri Institute for Pharmacological Research, Milan, Italy.* vittorio.berteletti@marionegri.it

Focus on more specific fetal testing

Most prenatal tests analyse a mix of maternal and placental DNA, and cannot distinguish the source of any abnormality (D. W. Bianchi *Nature* 522, 29–30; 2015). Our test, which was used to screen 185,000 women worldwide last year, reports specifically on fetal chromosomal status and so gets around the problem of validating incidental discoveries related to the mother's health.

The test detects single nucleotide polymorphisms (SNPs) in maternal and fetal DNA. The results are used to calculate a pregnancy-specific risk score; the maternal contribution is used only as input to the fetal-risk calculation (see B. Zimmermann *et al. Prenat. Diagn.* 32, 1233–1241; 2012). Therefore, incidental findings of maternal risk for cancer or abnormal numbers of sex chromosomes are not an issue. An exception is the partial deletion of chromosome 22 associated with DiGeorge syndrome. By providing the

fetal-risk score, the test may identify the mother as a carrier of this deletion, as our consent form mentions.

Without knowing the sensitivities, specificities and positive or negative predictive values of incidental maternal findings, their clinical significance and utility are unknown. For example, there is currently no way to estimate residual risk when tests are negative for maternal cancer.

Susan J. Gross, Stephanie Kareht, Allison Ryan Natera, *San Carlos, California, USA.* sgross@natera.com
Competing financial interests declared: see go.nature.com/kgd6ik.

Pollution hotspots tracked in real time

As Beijing citizens, we are glad that China is prioritizing pollution monitoring as it starts to implement its new environmental protection law. Advances in environmental monitoring technology — including wireless sensors, mobile apps and big-data processing — are raising public awareness and participation. These facilitate analysis by directly connecting researchers with government agencies and non-governmental organizations.

For example, wireless sensors installed on the chimney tops or drainage outlets of polluting factories monitor emissions such as sulfur dioxide and nitrogen oxides, or measure water contamination — all in real time. Big-data technology allows the rapid conversion of these massive data sets into regional and national insights that can be visualized online. For instance, real-time air-quality data from China's 367 major cities are mapped at www.cnemc.cn (in Chinese).

Xiaoyan Zheng, Yibing Lv *China National Environmental Monitoring Center, Beijing, China.* zhengxy@cnemc.cn

FORUM Developmental biology

Nanotubes in the niche

In fruit flies, protrusions can extend from stem cells in the testes to cells in a regulatory hub, mediating intercellular signalling and stem-cell maintenance. The implications of this finding are presented here from two angles. [SEE LETTER P.329](#)

THE PAPER IN BRIEF

- In the testes of fruit flies, germline stem cells, which give rise to sperm, divide asymmetrically.
- One daughter cell retains its stem-cell identity and remains attached to an adjacent cluster of cells called a hub, whereas the other daughter is displaced and differentiates.
- The hub secretes Dpp, a member of the BMP family of proteins, which signals to and

regulates maintenance of germline stem cells. But how this signalling is prevented from acting in differentiating daughters is unclear.

- On page 329 of this issue, Inaba *et al.*¹ demonstrate that germline stem cells form protrusions dubbed microtubule-based nanotubes, which extend to the hub and mediate signalling between Dpp in the hub and its receptor proteins in germline stem cells (Fig. 1).

San Francisco, California 94143, USA.
e-mail: tkornberg@ucsf.edu

Reach for self-renewal

LILACH GILBOA

To maintain tissue integrity, stem cells must strike a balance between two fates: self-renewal and differentiation. In the testes of fruit flies, this balance is determined by proximity to the hub — a source of the self-renewing factor Dpp⁹. Many have questioned how the spread of Dpp from the hub is limited, to create the sharp concentration gradient that distinguishes self-renewal from differentiation across one cell diameter. Inaba and colleagues' work suggests that the fate of GSC daughters is determined not by the extracellular spread of Dpp, as had been posited, but by the ability of these cells to directly access Dpp at its source. This change of view requires a general reconsideration of how niche components affect stem-cell self-renewal.

Inaba *et al.* report that Dpp receptors concentrate in puncta (clusters) on MT-nanotubes, which extend from GSCs into the hub. Crucially, increasing the thickness of MT-nanotubes increases both the number of Dpp-receptor puncta and the responsiveness of GSCs to Dpp. This result points to an unexpected property of the GSC self-renewal system — the amount of ligand available is not the limiting factor. Furthermore, GSCs can use more Dpp than normal only if they have MT-nanotubes of increased thickness. Conversely, when the authors shortened MT-nanotubes, Dpp pathway activity decreased in GSCs. The Dpp ligand, which presumably was no longer being sequestered by GSCs, was still unable to elicit Dpp signalling in daughters removed from the niche. These data are incompatible with a simple model in which cell fate is determined by a diffusible ligand, and instead suggest that MT-nanotubes constitute the main way in which Dpp is accessed.

Previous studies^{10–12} found a role for proteins

Close encounters

THOMAS B. KORNBERG

Development is not a democracy. As cells arrange themselves into complex patterns and structures, some act as directors and some as their clients. This principle was first uncovered more than 100 years ago in a study of regeneration in the freshwater animal *Hydra*², and was crafted into a general theory by work on amphibian development³. These early studies showed that not only can some cells (directors) remember their origins when transplanted to abnormal sites in the embryo, but they can also persuade neighbouring host cells (clients) to join them and make structures from that origin. This led to the idea that cell–cell signalling over short distances has a key role in development. How such paracrine signals are relayed has long been a subject of study — one that is addressed by Inaba *et al.* in the germline stem cells (GSCs) of the testes of fruit flies.

The GSC niche is an excellent environment for studying how paracrine signals move from directors to clients and how signals selectively act only on intended targets, because selective Dpp signalling to GSCs is imperative to ensure that some cells self-renew and some become sperm. Inaba and colleagues show that GSCs make microtubule-based nanotubes (MT-nanotubes) that extend into the hub, with which they seem to pick up Dpp. There is no Dpp signalling or self-renewal in GSCs with defective MT-nanotubes, indicating that these cells need to pick up Dpp directly. Therefore,

MT-nanotube-mediated protein exchange ensures that Dpp signals selectively to GSCs.

Such direct transfer of signalling proteins between cells is not restricted to GSCs — in fact, it might be a universal method of paracrine signalling. Research from several groups has indicated^{4–8} that paracrine signals can be transmitted through cellular protrusions called cytonemes. These structures, which are primarily composed of a structural protein called actin, are a specialized form of the group of cytoplasmic projections called filopodia. Cytonemes have been shown to transport a range of paracrine signalling proteins: Dpp, Hedgehog, fibroblast growth factor and Wntless proteins in fruit flies^{4–6}; Sonic hedgehog in the developing chick limb⁷; and Wnt protein in developing zebrafish embryos⁸. These processes of transport and exchange by cytonemes are similar to that reported in the current study.

Signal-mediating protrusions can be short or long, composed of actin filaments or microtubules, and might extend from director to client or from client to director. But in all contexts studied, exchange of signalling proteins occurs between the protrusion and its target cell. It seems that biology has created a varied set of structures to move signals between cells by this basic mechanism. The coming years promise to reveal how these structures select their targets and make functional contacts with them, and how they transport, release and take up signals.

Thomas B. Kornberg is in the Cardiovascular Research Institute, University of California, San Francisco,

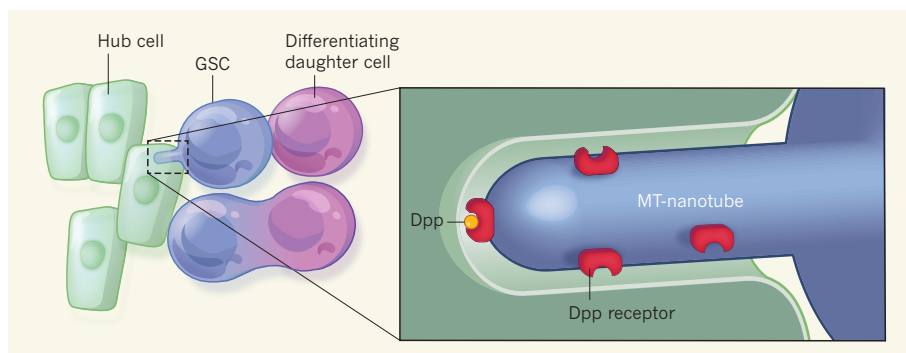


Figure 1 | Maintenance of germline stem cells. In the testes of fruit flies, germline stem cells (GSCs) reside in close proximity to a cellular hub, which produces the signalling protein Dpp. On cell division, the daughter closest to the hub retains its stem-cell identity owing to Dpp signalling. However, Dpp signalling is not activated in the other daughter, which subsequently differentiates to produce sperm. Inaba and colleagues¹ report that this selectivity is mediated by microtubule-based (MT) nanotubes that protrude from GSCs to pick up Dpp from cells of the hub. These protrusions contain clusters of Dpp receptor proteins, which transduce Dpp signalling in GSCs and so induce self-renewal. The right-hand box is a cartoon based on Figure 5e of Inaba and colleagues' paper; the details of the depicted process, including the topology of the nanotubes and localization of receptors, are not yet known.

in the extracellular matrix in maintaining GSCs, presumably by presenting the cells with maintenance factors. How should we view these findings in light of the current study? One explanation could be that components of the extracellular matrix stabilize MT-nanotubes, or promote Dpp–receptor interactions in some other way. Alternatively, extracellular-matrix components might affect other proteins secreted by the niche, such as Unpaired, which

supports GSC maintenance by promoting the cells' adhesion to the hub¹³. Future studies will determine whether this adhesion is a prerequisite for MT-nanotube formation, or whether the two pathways act independently.

MT-nanotubes, or similar structures, might well promote stem-cell maintenance in other organs and organisms. If this is the case, there must be a fundamental change in our efforts to understand stem-cell maintenance,

moving away from attempts to discover how extracellular ligand spread is limited to stem cells, and towards how stem cells access the ligands that are needed for self-renewal. As our understanding of the connection between the stem cell and its niche continues to increase, one thing is clear — the study of GSCs in fruit flies will continue to provide important insights. ■

Lilach Gilboa is in the Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel.
e-mail: lilacgilboa@weizmann.ac.il

1. Inaba, M., Buszczak, M. & Yamashita, Y. M. *Nature* **523**, 329–332 (2015).
2. Browne, N. E. *J. Exp. Zool.* **7**, 1–23 (1909).
3. Spemann, H. & Mangold, H. *Wilhelm Roux's Arch. Dev. Biol.* **109**, 599–638 (1924).
4. Huang, H. & Kornberg, T. B. *eLife* **4**, e06114 (2015).
5. Roy, S., Huang, H., Liu, S. & Kornberg, T. B. *Science* **343**, 1244–1246 (2014).
6. Bischoff, M. et al. *Nature Cell Biol.* **15**, 1269–1281 (2013).
7. Sanders, T. A., Llagostera, E. & Barna, M. *Nature* **497**, 628–632 (2013).
8. Stanganello, E. et al. *Nature Commun.* **6**, 5846 (2015).
9. Kawase, E., Wong, M. D., Ding, B. C. & Xie, T. *Development* **131**, 1365–1375 (2004).
10. Zheng, Q., Wang, Y., Vargas, E. & DiNardo, S. *Dev. Biol.* **357**, 202–210 (2011).
11. Guo, Z. & Wang, Z. *Development* **136**, 3627–3635 (2009).
12. Hayashi, Y., Kobayashi, S. & Nakato, H. *J. Cell Biol.* **187**, 473–480 (2009).
13. Leatherman, J. L. & Dinardo, S. *Nature Cell Biol.* **12**, 806–811 (2010).

This article was published online on 1 July 2015.

EARTH SCIENCE

Big geochemistry

A compilation of more than 300,000 rock compositions provides crucial input into a 100-year-old debate on how the continental crust formed, and provides new constraints for theories of continental-crust development. SEE ARTICLE P.301

CHRISTY TILL

Earth is the only planet in our Solar System known to have a buoyant continental crust. The origin of this crust has been long debated because its compositional heterogeneity precludes the straightforward testing of formation models. On page 301 of this issue, Keller *et al.*¹ present perhaps the most convincing data so far on this issue. They demonstrate that most of the continental crust's igneous rocks — those formed by the solidification of lava or magma — are formed through progressive crystallization of melted material extracted from Earth's interior, followed by the return of the most dense crystals to the mantle.

Two competing hypotheses that arose in the pioneering experimental geochemistry laboratories of the early 1900s have influenced nearly

all models for the formation of the continental crust. In 1915, Norman L. Bowen proposed that the continental crust formed through the progressive crystallization and distillation of magmas generated in Earth's mantle² (Fig. 1a). In this theory, the first minerals to crystallize are those with the highest abundance of iron and magnesium (mafic minerals), leaving the remaining melt enriched in silica (SiO₂). Mantle-derived magma that originally contains 50% silica can thus evolve through continuous crystallization to form a composition that matches that of the bulk continental crust (approximately 61% silica). In 1933, Bowen's mentor, Reginald A. Daly, proposed that assimilation was at least as important in forming the modern continental crust — ascending mantle-derived magmas drive the melting of the pre-existing continental crust, and mixing of the two, to produce the

observed silica content of the bulk continental crust³ (Fig. 1b).

As the problems and costs of geochemical analyses and computational time decrease, geoscientists can use ever-larger data sets to answer fundamental questions about our planet. Keller and colleagues¹ report one of the most compelling examples of this so far. They compiled more than 300,000 existing geochemical analyses of igneous rocks from around the world to calculate the average composition of magmas in areas where continental plates converge or are pulled apart (rifted).

By comparing the chemical compositions of plutonic rocks (which formed from magmas that slowly cooled within the Earth) with those of volcanic rocks (formed when magmas erupted on Earth's surface), the researchers assessed the dominant processes that contributed to the rocks' formation. If Daly's assimilation theory is correct, there should be a linear relationship between the concentration of oxides, such as magnesium oxide, and silica. However, Keller and co-workers observe a nonlinear relationship, which confirms the dominance of Bowen's crystallization hypothesis.

One long-standing issue with Bowen's hypothesis is that mafic crystal residues should have accumulated in the lower crust over time, forming cumulate rock. Several lines of

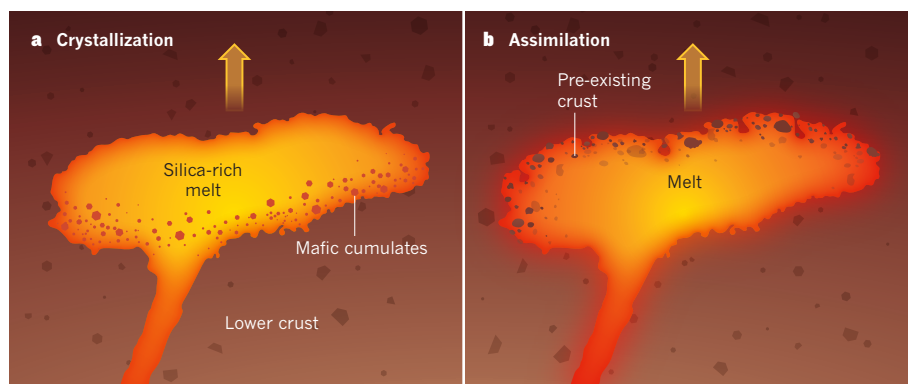


Figure 1 | Models of continental-crust formation. **a**, In the crystallization model, melted material within the mantle rises to the surface, inducing the crystallization of minerals from the melt as the temperature of the surrounding environment falls. Mafic minerals (those with the highest abundance of iron and magnesium) are the first to crystallize and fall to the bottom of the melt, where they form rocks known as mafic cumulates. The remaining melt is enriched in silica and forms the continental crust. **b**, In the assimilation model, ascending mantle-derived magmas drive the melting of the pre-existing continental crust, which mixes into the magma. The resulting mixture goes on to form new continental crust. Keller and colleagues' compilation¹ of geochemical data from around the globe supports the idea that crystallization is the dominant process in modern continental-crust formation.

evidence — including data from geological samples, seismic imaging, elemental mass-balance calculations and numerical simulations — support a model in which mafic crystal residues can reach a density greater than that of the uppermost mantle, causing them to detach from the base of the continental crust and sink back into the mantle, a process called delamination^{4–8}. Keller *et al.* identify key differences between the average plutonic and volcanic rock compositions from convergent and rift environments. These differences are consistent with predicted cumulate compositions, based on the authors' extensive calculations, and demonstrate that mafic-crystal cumulates are present in plutonic rocks, but not in large abundances. The authors' observation that most continental plutonic-rock compositions have higher silica content than expected for mafic cumulates substantiates both Bowen's hypothesis and the delamination model.

Keller *et al.* also conclude that plutonic rocks that formed in continental-rift environments exhibit characteristics consistent with the presence of water during their formation. Such environments were previously thought to be dry. The presence of water could explain why these magmas stalled and crystallized in the continental crust instead of erupting at the surface: water-containing magmas approach the temperature at which crystals start to form as they ascend, whereas dry magmas do not. Ultimately, the bulk continental crust has geochemical signatures consistent with formation in the presence of small amounts of water⁹. The observation that both convergent and rift plutonic environments produce such signatures, and that they preferentially crystallize in the crust rather than erupt, therefore provides new constraints on theories of how the continental crust developed.

Two notable questions relating to Keller and colleagues' findings are when continental-crust formation started, and whether crustal formation processes have varied over Earth's history — topics that two of the authors of this work wrestled with in an earlier paper¹⁰. Statistical evidence in the present data set suggests that the composition of the population of rocks more than 100 million years old does not vary substantially from that of the younger population. But the data set for older rocks is much smaller than that for younger ones, and many of the older rocks have been lost through recycling processes such as subduction and erosion; those that survived recycling may

not be representative of the diversity that once existed. The observations now presented by Keller *et al.* are therefore most applicable to modern continental-crust formation.

In an era in which breakthroughs in geochemistry are often attained by probing rocks and minerals at the nanometre and atomic scales¹¹, Keller *et al.* are pushing the boundaries in the other direction, reminding us of the value of investigations averaged over large scales of time and space. Since the pioneering days of Bowen and Daly, the field of geochemistry has evolved to incorporate observations ranging from the microscopic to the global in scale. Future work will require the integration of such geochemical observations with those from a wide range of other disciplines to explore questions such as whether the formation of the continental crust played a primary part in the rise of oxygen in Earth's atmosphere. ■

Christy Till is at the School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85281, USA.
e-mail: christy.till@asu.edu

1. Keller, C. B., Schoene, B., Barboni, M., Samperton, K. M. & Husson, J. M. *Nature* **523**, 301–307 (2015).
2. Bowen, N. L. *J. Geol.* **23** (suppl.), 1–91 (1915).
3. Daly, R. A. *Igneous Rocks and the Depths of the Earth* (McGraw-Hill, 1933).
4. Kay, R. W. & Kay, S. M. *Tectonophysics* **219**, 177–189 (1993).
5. Jull, M. & Kelemen, P. B. *J. Geophys. Res.* **106**, 6423–6446 (2001).
6. Gao, S. *et al.* *Nature* **432**, 892–897 (2004).
7. Plank, T. J. *Petrol.* **46**, 921–944 (2005).
8. Levander, A. *et al.* *Nature* **472**, 461–465 (2011).
9. Rudnick, R. L. *Nature* **378**, 571–578 (1995).
10. Keller, C. B. & Schoene, B. *Nature* **485**, 490–493 (2012).
11. Valley, J. W. *et al.* *Nature Geosci.* **7**, 219–223 (2014).

CANCER

A dendritic-cell brake on antitumour immunity

Activation of a cellular stress response and the transcription factor XBP1 in dendritic cells has now been shown to limit the cells' ability to stimulate antitumour immune responses in a mouse model of ovarian cancer.

MIRIAM MERAD & HÉLÈNE SALMON

Cells in the microenvironment of a tumour are exposed to stressful events. This is due in part to a lack of oxygen and nutrients that leads to malfunction of the endoplasmic reticulum (ER), the cellular organelle responsible for protein folding. The result is an accumulation of unfolded proteins — a state referred to as ER stress. Cancer cells use a signalling pathway known

as the ER stress response to survive in this environment. Whether this pathway also affects non-cancer cells in the tumour microenvironment, which are exposed to similar conditions, was not known. Writing in *Cell*, Cubillos-Ruiz *et al.*¹ show, in a model of ovarian cancer, that induction of the ER stress response in tumour-associated dendritic cells disrupts the cells' ability to promote adaptive antitumour immunity.

Dendritic cells (DCs) are a type of immune

cell found in all tissues. DCs are classed as antigen-presenting cells (APCs) because they present characteristic molecular structures (antigens) expressed by tissue cells to the immune system's T cells and so elicit tissue-specific T-cell responses². In the tumour environment, DCs sample, process and present tumour-cell-associated antigens and are thus potent inducers of T-cell responses that can eliminate target tumour cells. Other APCs include tissue-resident macrophages and circulating monocytes. Although these are less potent than DCs, they can also modulate T-cell effector responses².

The key role of T-cell responses in cancer treatment has been firmly established in recent years through the remarkable clinical success of strategies aimed at increasing the activation of tumour-targeting T cells. These therapies operate through the engineering of T cells to carry receptors that bind strongly to tumour antigens or by blocking inhibitory molecules on T cells to increase T-cell function. However, only subsets of patients respond to these treatments, and there is an urgent need to identify other brakes on antitumour immunity.

Although interactions between T cells and DCs typically occur in tissue-draining lymph nodes, it is now clear that DC presentation of tumour antigens also occurs in the tumour itself, suggesting that DCs in the tumour microenvironment can substantially influence the functions of antitumour T cells. Thus, it is crucial to assess the characteristics of tumour-associated DCs and to identify molecules for maximizing local antigen presentation that might also be potential drug targets. Altered DC function is common in tumours, and different mechanisms that underlie DC dysregulation in the tumour microenvironment have been identified, including reduced DC accumulation at the tumour site³ and increased immunosuppression induced by soluble factors (cytokines) that are produced by tumour cells or their surrounding tissue⁴.

The ER stress response is mediated by several signalling molecules, among which the enzyme IRE1 α and its target transcription factor XBP1 are the most evolutionarily conserved. XBP1 has been shown to promote tumour growth when expressed in cancer cells⁵. Using a mouse model of ovarian cancer, Cubillos-Ruiz *et al.* show that damaging molecules called reactive oxygen species (ROS) accumulate in tumour-associated DCs, and that ROS cause lipid degradation (by peroxidation) and the accumulation of lipid by-products, which in turn leads to the ER stress response and XBP1 activation (Fig. 1). Analysis of the genes expressed by these DCs revealed that a lipid-synthesis program is induced in XBP1-expressing cells but not in those lacking XBP1.

The authors went on to show that deletion of the gene that encodes XBP1 in cells expressing CD11c (an integrin protein expressed at high

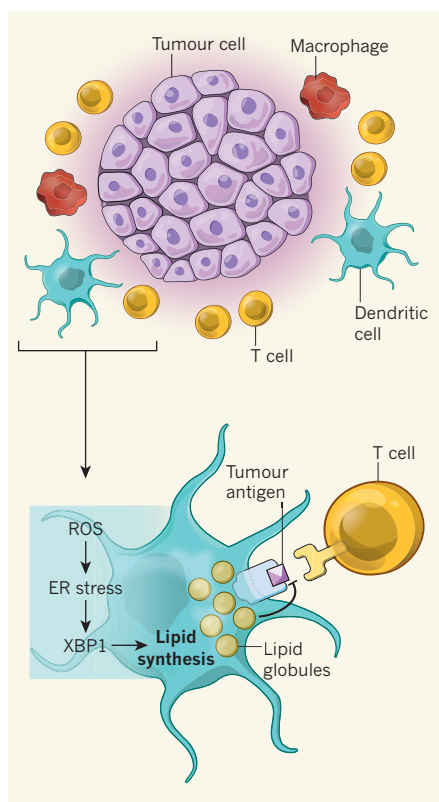


Figure 1 | XBP1 activation inhibits presentation of tumour antigens. Antitumour immune responses rely on dendritic cells (DCs) to present molecular structures (antigens) from tumour cells to T cells, which can then kill the antigen-bearing tumour cells. Cubillos-Ruiz *et al.*¹ show that damaging molecules called reactive oxygen species (ROS), which are produced in tumour-associated DCs, induce an endoplasmic reticulum (ER) stress response. The ER stress pathway leads to expression of the transcription factor XBP1, which induces lipid synthesis. The resulting accumulation of lipids in the DCs reduces their ability to present antigens and thus impairs antitumour T-cell responses. Macrophage cells may also contribute to the ER stress response.

levels on DCs), or delivery of short inhibitory RNA molecules that inhibit XBP1 expression, led to reduced lipid accumulation in DCs, increased DC antigen-presentation capacity, greater accumulation of T cells in the ovarian tumours and reduced tumour growth. Scavenger molecules that eliminate lipids and reduce ROS levels similarly improved DC antigen-presentation ability, suggesting that XBP1's function in enhancing tumour formation is at least partly due to the induction of lipid biosynthesis in DCs.

Although DCs express high levels of CD11c in the steady state, other APCs such as macrophages and monocyte-derived cells can also express CD11c (ref. 2), and so the therapeutic benefit of XBP1 deletion in CD11c-expressing (CD11c⁺) cells might not solely reflect its effect on DCs. XBP1 has an important role in the physiology of APCs under homeostatic conditions: deletion of XBP1 in non-tumour

DCs compromises their survival⁶ and antigen-presentation ability⁷, and loss of XBP1 in activated macrophages reduces cytokine release in response to microbial stimuli⁸. These findings contrast with Cubillos-Ruiz and colleagues' demonstration that XBP1 deletion in tumour-associated CD11c⁺ cells does not compromise DC survival and improves their antigen-presentation function. Such surprising results suggest that the pro-immunogenic homeostatic function of XBP1 is subverted in tumour-associated DCs. The extent to which the tumour or its environment promotes XBP1 activation or subverts its function should be explored further.

Increased lipid accumulation has been observed in tumour-associated CD11c⁺ cells in tumour-bearing mice and in patients with cancer⁹. Whether the lipid-laden CD11c⁺ cells described in these studies represent bona fide DCs, monocyte-derived cells or macrophages will need to be addressed by cell-specific assays. Lipid accumulation in macrophages has also been linked to neurodegenerative diseases and atherosclerosis, and it will be interesting to explore whether XBP1 contributes to the progression of these conditions.

Cubillos-Ruiz and colleagues' results extend previous studies^{10,11} showing that metabolic changes, for example increased breakdown of the amino acids arginine and tryptophan in tumour-associated DCs, alter T-cell effector function. Greater understanding of the extent of metabolic alterations in tumour-associated DCs should help to identify yet other pathways of immune dysregulation in tumours. In the meantime, the authors' finding that XBP1 deletion in DCs increases T-cell infiltration of tumours and reduces tumour growth is a timely addition to the pressing search for targets that increase antitumour T-cell function. ■

Miriam Merad and Hélène Salmon are at the Tisch Cancer and Immunology Institute, Oncological Science Department, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.
e-mail: miriam.merad@mssm.edu

1. Cubillos-Ruiz, J. R. *et al.* *Cell* **161**, 1527–1538 (2015).
2. Merad, M., Sathe, P., Helft, J., Miller, J. & Mortha, A. *Annu. Rev. Immunol.* **31**, 563–604 (2013).
3. Spranger, S., Bao, R. & Gajewski, T. F. *Nature* <http://dx.doi.org/10.1038/nature14404> (2015).
4. Chaput, N., Conforti, R., Viaud, S., Spatz, A. & Zitvogel, L. *Oncogene* **27**, 5920–5931 (2008).
5. Chen, X. *et al.* *Nature* **508**, 103–107 (2014).
6. Iwakoshi, N. N., Pypaert, M. & Glimcher, L. H. *J. Exp. Med.* **204**, 2267–2275 (2007).
7. Osorio, F. *et al.* *Nature Immunol.* **15**, 248–257 (2014).
8. Martinon, F., Chen, X., Lee, A.-H. & Glimcher, L. H. *Nature Immunol.* **11**, 411–418 (2010).
9. Herber, D. L. *et al.* *Nature Med.* **16**, 880–886 (2010).
10. Norian, L. A. *et al.* *Cancer Res.* **69**, 3086–3094 (2009).
11. Mellor, A. L. & Munn, D. H. *Nature Rev. Immunol.* **4**, 762–774 (2004).

ASTROCHEMISTRY

Fullerene solves an interstellar puzzle

Laboratory measurements confirm that a ‘buckyball’ ion is responsible for two near-infrared absorption features found in spectra of the interstellar medium, casting light on a century-old astrochemical mystery. [SEE LETTER P.322](#)

PASCALE EHRENFREUND & BERNARD FOING

More than 400 unidentified absorption features have been recorded in spectra of the interstellar medium of the Milky Way and other galaxies. The nature of these ‘diffuse interstellar bands’, which are present in spectra of the interstellar medium from the near-ultraviolet to the near-infrared domains, has evaded elucidation since their discovery nearly 100 years ago — one of the longest-standing mysteries of modern astronomy. Writing on page 322 of this issue, Campbell *et al.*¹ report laboratory spectra of the buckminsterfullerene, or ‘buckyball’, ion C_{60}^+ . The spectra provide an excellent match to two interstellar bands that we discovered and attributed to this ion², and securely identify C_{60}^+ as a component of the interstellar medium.

During the past 40 years, advances in astronomical instrumentation have enabled the detection of many weak diffuse interstellar bands (DIBs) (Fig. 1), and their various widths have been accurately measured^{3–5}. Numerous efforts have been made to identify the likely molecular group (the ‘carrier’) that causes the absorption bands through laboratory

studies, theoretical modelling and astronomical observations⁴. As a result, a host of candidate carriers has been proposed, ranging from dust grains to intriguing carbonaceous species, including fullerenes (a carbon allotrope)^{3,6}. High-resolution observations of fine spectroscopic details of certain DIBs^{7,8} have contributed to the consensus that polyatomic carbonaceous molecules in the gas phase are promising candidate carriers. The discovery and synthesis of fullerenes (including buckminsterfullerene) and of graphene in laboratories, have revolutionized our understanding of carbon chemistry and its applications ranging from astronomy and biotechnology to industrial processes.

In 1994, a search for the strongest signals of C_{60}^+ in interstellar-material spectra led to the discovery² of two absorption features (Fig. 1) at 9,577 and 9,632 ångströms. These coincided with independent measurements of electronic transitions of C_{60}^+ in laboratory spectra using frozen-matrix techniques — in which the target species is diluted with an unreactive host species (such as neon) at very low temperatures, and is then spectrally analysed⁹. This preliminary match of two interstellar DIBs with C_{60}^+ showed wavelength shifts of 3–10 Å relative to the laboratory measurements; these

were attributed to known effects caused by the neon matrix. Nonetheless, the interstellar-medium study² estimated that about 0.6% of interstellar carbon may be in the form of C_{60}^+ , and proposed that fullerenes may have a significant role in interstellar chemistry. Arguments to support the C_{60}^+ match to the two strongest near-infrared DIBs also include the species’ extreme resistance to ultraviolet radiation, that the strengths and widths of the two DIBs were mutually correlated, and that their band widths were consistent with the spectrum of a rotating fullerene-like molecule^{2,10,11}. To provide unambiguous confirmation, laboratory gas-phase spectra of C_{60}^+ were called for.

In the current study, Campbell *et al.* report the laboratory synthesis of a complex of C_{60}^+ and helium at 5.8 kelvin, providing the best approximation so far of the gas-phase spectrum of C_{60}^+ . The laboratory measurements of C_{60}^+ band maxima at 9,632.7 and 9,577.5 Å show a stunning match to the observed wavelengths of the corresponding interstellar DIBs, with an accuracy of ± 0.1 Å. Moreover, the respective band widths of 2.2 and 2.5 Å are consistent with the 2.8-Å widths measured for these DIBs^{10,11}. Twenty-one years after the initial assignment, the authors’ superb laboratory results, which were achieved with a technically advanced set-up, have conclusively identified C_{60}^+ as the carrier of these two DIBs.

With the two strongest C_{60}^+ bands securely matched to DIBs, prospects are improving for the identification of the carriers of the more than 400 remaining DIB features, which have been recorded in a variety of astrophysical settings. In the past decade, observations across the entire optical range have established that diffuse band carriers are ubiquitous in space. DIBs have also recently been detected and mapped in the Magellanic Clouds (the Milky Way’s satellite galaxies) and in other galaxies. The carrier molecules may therefore constitute an important reservoir of organic material throughout the Universe.

The clear identification of additional DIB carriers could have a far-reaching impact on our understanding of astrochemical reaction networks. C_{60}^+ was possibly the easiest and most conspicuous molecule to identify owing to its signature of two strong bands in the near-infrared range that is clearly evident in interstellar-medium spectra, which have been corrected for atmospheric absorptions^{2,10,11}. The species’ cage-like symmetry and geometry⁶ (Fig. 1), as well as its chemical properties, increase its chances of survival in space and its expected natural formation in astrophysical environments^{4,10}. One possible approach to identify more DIB carriers would be to detect and quantify the presence in space of other fullerene compounds, which could be formed by the addition of hydrogen atoms, metals or other radicals in different charge or reactive states.

In previous analyses of mid-infrared

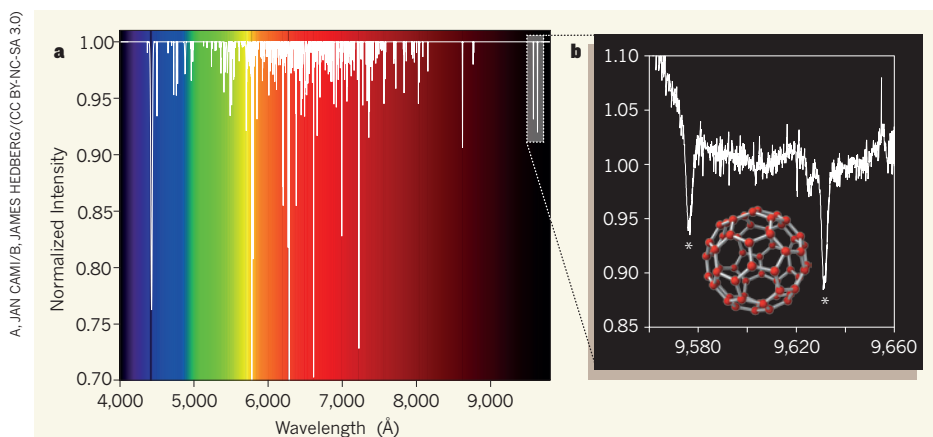


Figure 1 | Buckminsterfullerene in space. **a**, High-resolution absorption spectra of the interstellar medium, such as that shown here, include more than 400 unidentified diffuse interstellar bands (DIBs, white lines) spread across the optical and near-infrared domains^{3,5,10}. Fullerenes and other carbon-bearing molecules have been proposed as the absorbing species causing some of the interstellar DIBs, but conclusive identification had been elusive. **b**, Campbell *et al.*¹ now prove that the charged buckminsterfullerene ion C_{60}^+ is indeed responsible for the two strongest interstellar DIBs¹⁰ in the near-infrared at 9,577 and 9,632 Å (indicated by stars) discovered² in 1994.

emission spectra of a young planetary nebula (the ionized, ejected envelope of an old star) taken with NASA's Spitzer Space Telescope, researchers for the first time identified¹² specific molecular-vibration bands that were assigned to compounds of the fullerenes C_{60} and C_{70} . Vibrational spectroscopy cannot perfectly distinguish between molecules of similar chemical bonds or between molecular excitation states, but it can provide a global inventory of molecular groups from the pool of organic materials in space. Electronic spectroscopy, such as that used by Campbell and colleagues, can pinpoint the fingerprints of individual molecules. Additional observations of circumstellar nebulae have shown that, under the right conditions, fullerenes can and do form efficiently in space⁴.

Other carbon-bearing molecules known as polycyclic aromatic hydrocarbons (PAHs) — fused benzene rings — have also been proposed^{13,14} as carriers of vibrational bands that are observed in the mid-infrared domain. These are estimated to be the most abundant complex organic molecules in space, comprising more than about 15% of cosmic carbon, and have also been proposed as potential DIB carriers. However, in contrast to C_{60}^+ , PAHs can undergo partial loss of hydrogen atoms (dehydrogenation), and would therefore exist in a large number of molecular isomers, diluting their spectral signatures¹⁵. This might explain why, despite several dedicated efforts¹⁶, no specific PAH electronic feature has been clearly associated with DIBs. Some metal-PAH complexes may eventually be shown to have special molecular configurations that have few isomers and strong oscillator strengths — these would have undiluted spectral features, which would permit their electronic spectroscopic identification and matching to astrophysical DIBs⁴.

The relationships between PAHs and fullerenes in space represent an intriguing puzzle. Laboratory studies have shown¹⁷ that C_{60}^+ can form following sequential light-induced fragmentation of large PAH cations. Infrared observations obtained by the Spitzer and the European Space Agency's Herschel Space Observatory have also shown¹⁸ that C_{60} can form in the tenuous and cold environment of an interstellar cloud illuminated by strong ultraviolet starlight — possibly as the product of the conversion of PAHs to graphene, and then to fullerenes.

As Campbell and colleagues' findings show, the identification of space molecules can be achieved only by measuring laboratory spectra under astrophysically relevant conditions, and the spectra must unambiguously match the astronomical data. This task requires theoretical modelling to determine other plausible fullerene- and PAH-candidate DIB carriers and a systematic programme of laboratory spectroscopy of these in frozen neon matrices (for screening) and in the gas phase (for identification).

Astronomical surveys of DIB strengths and profiles in different interstellar and circumstellar environments are needed to elucidate the formation and fate of carriers. All these research disciplines must be combined to provide further insights into the photochemistry, thermal conditions and integration in protoplanetary-disk material that affect the inventory and evolution of organic molecules from molecular clouds to star- and planet-forming regions. ■

Pascale Ehrenfreund is at the Space Policy Institute, George Washington University, Washington DC 20052, USA. **Bernard Foing** is at the European Space Research and Technology Centre of the European Space Agency, ESA-ESTEC, 2200 Noordwijk, the Netherlands.
e-mail: pehren@gwu.edu; bernard.foing@esa.int

1. Campbell, E. K., Holz, M., Gerlich, D. & Maier, J. P. *Nature* **523**, 322–328 (2015).
2. Foing, B. H. & Ehrenfreund, P. *Nature* **369**, 296 (1994).

3. Herbig, G. H. *Annu. Rev. Astron. Astrophys.* **33**, 19–73 (1985).
4. Cox, N. L. J. & Cami, J. (eds) *Proc. Int. Astron. Un. Vol. 9* (2014).
5. Hobbs, L. M. *et al. Astrophys. J.* **705**, 32–45 (2009).
6. Kroto, H. W. *et al. Nature* **318**, 162–163 (1985).
7. Ehrenfreund, P. & Foing, B. H. *Astron. Astrophys.* **307**, L25–L28 (1996).
8. Sarre, P. *et al. Mon. Not. R. Astron. Soc.* **277**, L41–L43 (1995).
9. Fulara, J., Jakobi, M. & Maier, J. P. *Chem. Phys. Lett.* **211**, 227–234 (1993).
10. Foing, B. H. & Ehrenfreund, P. *Astron. Astrophys.* **317**, L59–L62 (1997).
11. Galazutdinov, G. A. *et al. Mon. Not. R. Astron. Soc.* **317**, 750–758 (2000).
12. Cami, J., Bernard-Salas, J., Peeters, E. & Malek, S. E. *Science* **329**, 1180–1182 (2010).
13. Leger, A. & Puget, J. L. *Astron. Astrophys.* **137**, L5–L8 (1984).
14. Allamandola, L. J., Tielens, A. G. G. M. & Barker, J. R. *Astrophys. J.* **290**, L25–L28 (1985).
15. Vuong, M. H. & Foing, B. H. *Astron. Astrophys.* **363**, L5–L8 (2000).
16. Salama, F. *et al. Astrophys. J.* **728**, 154–162 (2011).
17. Zhen, J., Castellanos, P., Paardekooper, D. M., Linnartz, H. & Tielens, A. G. G. M. *Astrophys. J.* **797**, L30–L34 (2014).
18. Berné, O. & Tielens, A. G. G. M. *Proc. Natl Acad. Sci. USA* **109**, 401–406 (2012).

SYSTEMS BIOLOGY

Network evolution hinges on history

The effects of mutations in proteins can depend on the occurrence of previous mutations. It emerges that such historical contingency is also important during the evolution of gene regulatory networks. SEE LETTER P.361

AARON M. NEW & BEN LEHNER

Anyone who has dealt with bureaucracy knows that the order in which events occur can be vital to achieving a given task. This is also true for evolution, in which certain mutations are sometimes beneficial, or even viable, only if other mutations have occurred first¹ — a phenomenon known as historical contingency. In this issue, Sorrells *et al.*² (page 361) show how historical contingency constrained the evolution of an entire gene regulatory network.

Historical contingency has long been thought³ to have a role in evolution, but only since the advent of large-scale DNA sequencing have we been able to peer back in time and replay the evolutionary tape to establish how the process works in biological systems. A wealth of DNA-sequence information means that we can now infer the sequences of extinct genes, allowing us to resurrect the proteins that they encoded⁴. Studying the role of these ancestral proteins in living cells has revealed that only certain paths can be taken during the evolution of protein functions⁵. But the part played by historical contingency in the

evolution of gene regulatory networks has not been well studied.

Budding yeasts, a group of fungi that includes the model organism *Saccharomyces cerevisiae* (brewer's yeast), have two mating types, 'a' and 'α'. When exposed to mating pheromone, both mating types switch on genes necessary for mating, and a cells also switch on genes specifically required for mating with a cells, referred to as a-specific genes (asgs). The identity of these asgs has changed little over hundreds of millions of years of evolution^{2,6,7}, but the way in which their expression is controlled has altered considerably.

Sorrells *et al.* reconstructed the evolution of asg regulation in *S. cerevisiae*. They showed that a regulatory protein, Ste12, initially activated asgs indirectly by binding to a DNA-binding protein called a2, and then evolved to bind directly to DNA. By constructing plausible evolutionary intermediates, the authors provide evidence that this change in Ste12 function was made possible by another change, in the way the two mating types are defined.

Previous work from this group^{6,7} found that, in ancestral species, a cells were defined by the presence of a2. But in *S. cerevisiae*, a cells are

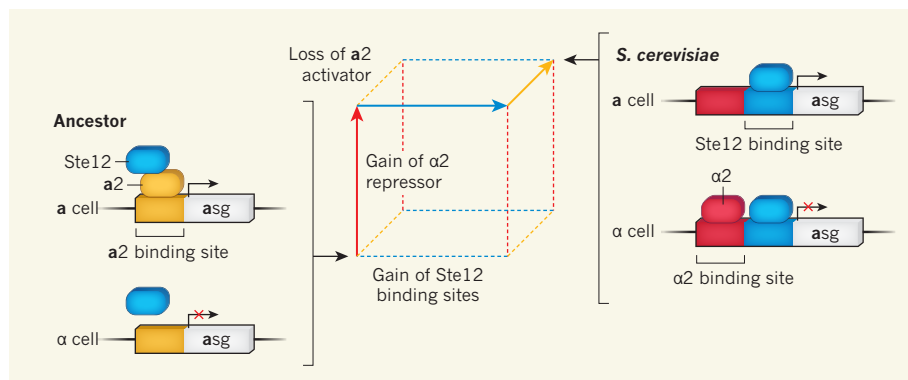


Figure 1 | Not all evolutionary paths are possible. Yeast cells exist in two mating types, *a* and α . In the ancestor of *Saccharomyces cerevisiae*, a regulatory protein called Ste12 activated *a*-specific genes (*asgs*) in *a* cells indirectly, by binding to a transcriptional activator protein called *a2* that was expressed only in *a* cells. By contrast, in present-day *S. cerevisiae*, Ste12 binds directly to *asgs*, and a transcriptional repressor protein called $\alpha 2$ represses the action of Ste12 in α cells. This evolution required three stages of genetic change, indicated on the red, blue and yellow axes. Sorrells *et al.*² demonstrate that, although there are six possible ways in which ancestral yeast could theoretically have evolved into the present-day yeast, only one is feasible (coloured arrows).

defined by the absence of a transcriptional repressor protein, $\alpha 2$, which is expressed only in α cells. Sorrells and colleagues argue that the evolution of direct binding sites for Ste12 could not have occurred before the evolution of $\alpha 2$, because $\alpha 2$ counteracts the effects of direct Ste12 binding in α cells — if $\alpha 2$ is absent, α cells misexpress *asgs*.

The authors propose that the loss of $\alpha 2$, which became possible because of the presence of $\alpha 2$, must have been the final step, making indirect regulation of *asgs* by Ste12 impossible. As such, although there are six orders in which these three evolutionary events could theoretically have occurred, only one is actually possible: gain of $\alpha 2$, then direct regulation of *asgs* by Ste12, followed by loss of $\alpha 2$ (Fig. 1).

Sorrells and colleagues' work illustrates that it can be naive to try to understand the evolution of individual genes without considering the evolution of the regulatory networks in which they act. Similarly, it can be a folly to study the evolution of isolated networks and functions without the broader context. In this case, a change in the way cells define their mating type was required for a change in how they switch on genes when exposed to mating pheromone. Even in single-celled organisms, genes and networks are typically reused to achieve many different purposes. The mechanism by which one process is achieved can therefore affect how another can evolve. In multicellular organisms, the situation is likely to be even more complex, as genes and networks are reused in different cell types and organs.

It is worth noting that Sorrells *et al.* treated the two regulatory networks they studied as simple linear hierarchies of interactions. But much biological regulation, including that of the mating-pheromone response in yeast, involves feedback and complex nonlinear dynamics^{8,9}. In future studies, it will be necessary to investigate how these nonlinear

interactions between genes and proteins further constrain or facilitate the evolution of molecules and networks¹⁰.

Finally, this study exemplifies the power of using yeast as a system to understand evolution through the reconstruction and experimental characterization of ancestral genes and their interactions. The phylogeny of yeast provides an unusual situation, because genetic-engineering tools developed to modulate brewer's yeast can be applied across several

species — two experiments in this study were performed in 'non-model' species that have evolutionary intermediate-like mechanisms of regulation. Molecular evolution is sometimes criticized for not allowing the rigorous experimental testing of hypotheses, but the ability to functionally characterize evolutionary intermediates means that this is no longer the case. ■

Aaron M. New and Ben Lehner are at the EMBL-CRG Systems Biology Unit, Centre for Genomic Regulation (CRG), 08003 Barcelona, and at the Universitat Pompeu Fabra, Barcelona, Spain. B.L. is also at the Institut Catalana de Recerca i Estudis Avançats, Barcelona.

e-mail: ben.lehner@crg.eu

1. Lehner, B. *Trends Genet.* **27**, 323–331 (2011).
2. Sorrells, T. R., Booth, L. N., Tuch, B. B. & Johnson, A. D. *Nature* **523**, 361–365 (2015).
3. Wright, S. *Proc. VI Int. Congr. Genet.* Vol 1, 356–366 (1932).
4. Harms, M. J. & Thornton, J. W. *Nature Rev. Genet.* **14**, 559–571 (2013).
5. DePristo, M. A., Weinreich, D. M. & Hartl, D. L. *Nature Rev. Genet.* **6**, 678–687 (2005).
6. Tsong, A. E., Tuch, B. B., Li, H. & Johnson, A. D. *Nature* **443**, 415–420 (2006).
7. Baker, C. R., Booth, L. N., Sorrells, T. R. & Johnson, A. D. *Cell* **151**, 80–95 (2012).
8. Yu, R. C. *et al.* *Nature* **456**, 755–761 (2008).
9. Tyson, J. J., Chen, K. C. & Novak, B. *Curr. Opin. Cell Biol.* **15**, 221–231 (2003).
10. Jiménez, A., Munteanu, A. & Sharpe, J. *Proc. Natl Acad. Sci. USA* **112**, 2103–2108 (2015).

This article was published online on 8 July 2015.

SYMBIOSIS

Receptive to infection

EPR3, a plant protein, is found to act as a probable receptor for exopolysaccharide molecules that surround the plant's symbiotic bacteria. The advance sheds light on how recognition is governed in symbiotic relationships. [SEE ARTICLE P.308](#)

SHARON R. LONG

The protective layers of polysaccharide molecules that bacteria build around themselves have varied roles, from buffering against environmental insults to disguising the cells from their hosts. These exopolysaccharides (EPS) are also thought to be involved in establishing symbiotic relationships, such as those between legume host plants and bacteria of the genus *Rhizobium* and related groups. On page 308 of this issue, Kawaharada *et al.*¹ move the decades-old quest to define this role for EPS a step forward, identifying a plant receptor protein that interacts with rhizobial EPS in a structurally specific manner.

In rhizobium–legume symbioses, bacteria induce plant roots to form specialized organs

called nodules. The rhizobia infect the nodule tissue and differentiate to form bacteroids, which convert nitrogen into ammonia that the plant can use to synthesize protein. The bacteria in turn benefit from plant-derived sugars. These relationships are often specific: particular strains or species of rhizobium typically form nitrogen-fixing root nodules on only a few species of legume². Such specificity is particularly interesting because of its evolutionary and ecological implications and its mechanistic intricacy — how do plants allow infection by a symbiotic friend but not a pathogenic foe?

According to a theory put forward in the 1970s, EPS mediate specificity by recognizing seed lectins (carbohydrate-binding proteins) of the corresponding host plant. However, this proposed role for lectins was not supported by

later genetic and biochemical studies. The first evidence against the theory was the finding³ that bacterial *nod* genes, which are key determinants of nodule formation, were expressed only when rhizobia obtained a signal from the plant and then made a signal in return — a crucial molecular conversation between bacteria and plants that had previously been missed.

The enzymes encoded by *nod* genes produce and modify small carbohydrates called Nod factors. Surprisingly, Nod factors are structurally quite different from EPS and other bacterial polysaccharides. They have a short chitin backbone, with side-chain modifications that vary for different bacterial species, such that they cause nodule formation only on certain plant hosts⁴. Nod factors interact with plant proteins, including Nod-factor receptor 1 (NFR1), NFR5 and SymRK, that act as receptors⁵. The signalling pathway activated by these Nod-factor receptors stimulates many plant responses, including nodule development and transcription⁶.

The discovery of Nod factors did not end the specificity saga⁷. In the 1980s, it emerged that EPS are essential for the actual invasion of plants^{8,9}, and several studies have suggested that the molecules' structures confer some infective specificity. For instance, the rhizobium *Mesorhizobium loti* produces EPS based on '8-sugar' monomers of ribose, glucose and galactose, and can make normal nodules on its host plant, *Lotus japonicus*. But a mutant form of *M. loti* that makes truncated EPS induces the formation of only small, uninfected nodule-like structures¹⁰. Kawaharada *et al.* reasoned that, although normal hosts can interact only with full-length EPS, a mutated form of the host receptor protein might interact with

truncated EPS, allowing normal nodulation. In a genetic 'suppressor' screen, they uncovered a mutant plant strain, *exo277*, that forms normal infected nodules with the mutant bacteria. The authors found that *exo277* harbours a single-nucleotide mutation in a gene that they dubbed *exopolysaccharide receptor 3* (*Epr3*).

Kawaharada and colleagues analysed the DNA sequence of wild-type *Epr3*, and found that the encoded protein is likely to be a membrane-spanning receptor-like protein showing striking similarity to NFR1. Both contain three extracellular LysM domains — regions that are closely related to chitin-binding proteins and are proposed to bind chitin-based molecules such as Nod factors. They also harbour an intracellular kinase domain, which might transduce the signal from the activated receptor into an intracellular signalling cascade. However, the amino-acid sequences around two EPR3 LysM domains differ from those of Nod-factor receptors: specific sequences normally found in evolutionarily conserved positions are located elsewhere, and are predicted to form different secondary structures. That the LysM domains are nonetheless important is shown by the fact that they are altered in several EPR3 mutants that show abnormal symbiotic behaviours. Thus, there is a correlation between altered LysM domains and changed receptor function.

Consistent with its predicted structure and function, the authors found that EPR3 is located at the cell membrane if artificially expressed in leaves of the plant *Nicotiana benthamii* and that, when purified, the extracellular domain binds to 8-sugar monomeric units of EPS. Finally, Kawaharada *et al.* demonstrated that *Epr3* gene expression

depends on Nod-factor signalling. Thus, early interactions between rhizobia and their plant partners involve two successive molecular conversations: first, the induction of Nod-factor synthesis in rhizobia and subsequent Nod-factor signal transduction in plant cells, leading to several plant responses, including transcription of EPS-receptor genes; and second, the perception of structurally specific EPS by those newly produced receptors (Fig. 1).

Kawaharada and colleagues' identification of a probable EPS receptor should open the door to the discovery of other components of the infection process. For instance, NFR1 and NFR5 are thought to act in concert with the co-receptor protein SymRK, which has an intracellular kinase domain and a distinct, leucine-rich-repeat type extracellular domain. One therefore wonders whether there might also be co-receptors for EPR3.

Furthermore, the findings highlight new areas of uncertainty that can be addressed experimentally. With what affinity and specificity does EPR3 bind EPS? How do LysM domains bind the non-chitinous EPS backbone? Efforts should also be directed to understanding how EPS binding acts to promote infection. Various mechanisms have been proposed, such as triggering positive plant responses that support bacterial activity, or downregulating plant defences; these hypotheses are now open to direct study.

Progress in *Lotus* plants should lead to advances in our understanding of rhizobial infection in other legumes, and ultimately enable us to identify conserved features that might be involved in pathogen–plant interactions as well as in other symbioses. Finally, identification of plant receptors required for rhizobial infection could lead to a more detailed understanding of basic plant cell biology, and of how plant defences can be overcome during bacterial invasion. ■

Sharon R. Long is in the Department of Biology, Stanford University, Stanford, California 94305-5020, USA.
e-mail: srl@stanford.edu

1. Kawaharada, Y. *et al.* *Nature* **523**, 308–312 (2015).
2. Brewin, N. J. *Annu. Rev. Cell Biol.* **7**, 191–226 (1991).
3. Fisher, R. F. & Long, S. R. *Nature* **357**, 655–660 (1992).
4. Dénarié, J., Debellé, F. & Promé, J.-C. *Annu. Rev. Biochem.* **65**, 503–535 (1996).
5. Gough, C. & Cullimore, J. *Mol. Plant–Microbe Interact.* **24**, 867–878 (2011).
6. Oldroyd, G. E. D. & Downie, J. A. *Annu. Rev. Plant Biol.* **59**, 519–546 (2008).
7. Gage, D. J. *Microbiol. Mol. Biol. Rev.* **68**, 280–300 (2004).
8. Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E. & Walker, G. C. *Nature Rev. Microbiol.* **5**, 619–633 (2007).
9. Niehaus, K. & Becker, A. in *Subcellular Biochemistry* Vol. 29 (eds Biswas, B. B. & Das, H. K.) 73–116 (Plenum, 1998).
10. Kelly, S. J. *et al.* *Mol. Plant–Microbe Interact.* **26**, 319–329 (2013).

This article was published online on 8 July 2015.

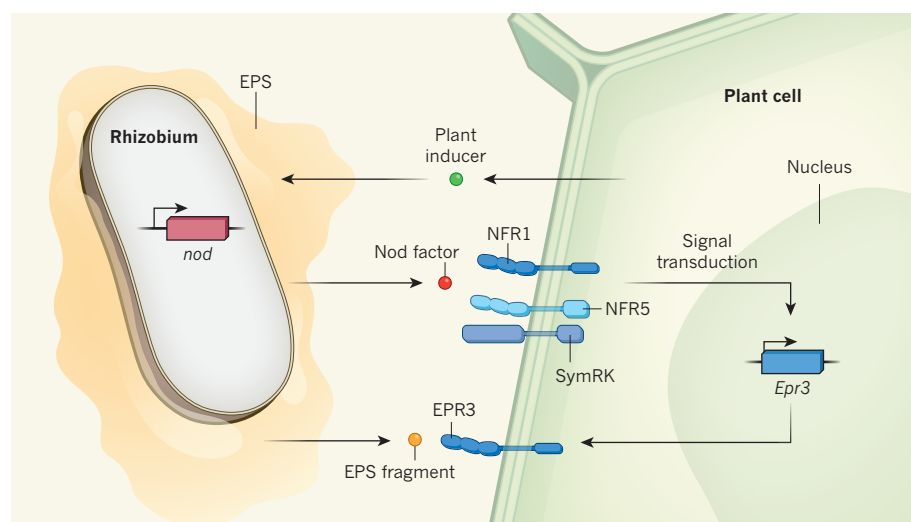


Figure 1 | Sequential steps in symbiosis. Bacteria of the genus *Rhizobium*, which are surrounded by a layer of exopolysaccharides (EPS), can have a symbiotic relationship with legumes. In the first stages of symbiosis, plant signals induce the transcription of rhizobial *nod* genes, which encode enzymes that synthesize Nod-factor (NF) proteins. NFs signal to the plant through NF receptor proteins, including NFR1, NFR5 and SymRK. Transduction of this signal results in several host responses that lead to the formation of root nodules, which the rhizobia invade and occupy as endosymbionts. In particular, Kawaharada *et al.*¹ report that NF signal transduction induces transcription of the gene *Epr3*, which they predict encodes a receptor protein that is intriguingly similar to NFR1-family receptors, and that can recognize EPS.

Volcanic–plutonic parity and the differentiation of the continental crust

C. Brenhin Keller¹, Blair Schoene¹, Melanie Barboni^{1,2}, Kyle M. Samperton¹ & Jon M. Husson^{1,3}

The continental crust is central to the biological and geological history of Earth. However, crustal heterogeneity has prevented a thorough geochemical comparison of its primary igneous building blocks—volcanic and plutonic rocks—and the processes by which they differentiate to felsic compositions. Our analysis of a comprehensive global data set of volcanic and plutonic whole-rock geochemistry shows that differentiation trends from primitive basaltic to felsic compositions for volcanic versus plutonic samples are generally indistinguishable in subduction-zone settings, but are divergent in continental rifts. Offsets in major- and trace-element differentiation patterns in rift settings suggest higher water content in plutonic magmas and reduced eruptibility of hydrous silicate magmas relative to dry rift volcanics. In both tectonic settings, our results indicate that fractional crystallization, rather than crustal melting, is predominantly responsible for the production of intermediate and felsic magmas, emphasizing the role of mafic cumulates as a residue of crustal differentiation.

Earth is the only planet in our Solar System with a high-silica continental crust, which is complemented by a low-silica oceanic crust similar to that of the Moon, Mars and Venus^{1,2}. The resulting bimodal topography³ of oceans and continents is essential for regulating Earth's climate^{4,5} and providing elemental nutrients to the biosphere⁶. Igneous geochemistry drives our understanding of the formation and evolution of the continental crust, along with its depleted complement, the mantle^{2,7}. Detailed investigations of specific igneous systems have revealed that mantle-derived basaltic melts differentiate to higher silica contents through processes such as fractional crystallization, metamorphism, and remelting of crustal material. Though these processes are generally agreed to result in a geochemically-stratified lithosphere which increases in silica abundance from the crust–mantle boundary to the surface^{2,8}, the compositional and geological heterogeneity of the continents—scaling from kilometres to micrometres—has made it difficult to assess the relative importance of specific endmember processes. Solving this problem requires an understanding of why magmas (1) ascend through the crust and erupt to form volcanic rocks, or (2) stall and freeze at depth as plutons. Some models suggest that magma composition and volatile content dictate eruptibility^{9,10}, while others emphasize physical processes such as crystal/melt segregation in magma chambers¹¹ resulting in complementary volcanic and plutonic compositions. Alternatively, eruptibility may be controlled by thermal constraints whereby higher magma fluxes result in more magma reaching the surface¹², in which case systematic geochemical differences between volcanic and plutonic rocks are not required. Here we test such models by taking advantage of increased availability of whole-rock compositional data to perform a comprehensive statistical analysis of the geochemical relationship between volcanic and plutonic rocks. We posit that this comparison provides insight not only into the predominant mechanisms of igneous differentiation and the origins of crustal stratification, but also—since magmas that reach the Earth's surface are more efficiently recycled back into the mantle via erosion and subduction—into crust–mantle geochemical evolution over billion-year timescales.

Quantifying volcanic–plutonic parity

To determine if continental plutonic and volcanic rocks are compositionally equivalent, we have applied computational statistical techniques to a data set of major- and trace-element geochemical analyses of 122,751 plutonic and 171,690 volcanic whole-rock samples of the continental crust, extracted from the EarthChem database¹³. After filtering spurious data (Methods), each data set was subjected to weighted bootstrap Monte Carlo analysis¹⁴, with sample weights inversely proportional to spatial sample density in order to obtain a uniform, maximally representative posterior sample distribution on the globe (Methods, Extended Data Fig. 1). This method maximizes the accuracy of statistical parameters such as the mean and median, and minimizes the effect of oversampling in regions that have been subjected to more intensive field study. Sampling bias is further minimized in our interpretations by focusing on intensive compositional variables, binned by differentiation proxies such as SiO₂ and MgO (Methods, Figs 1, 2). Results are reported as the mean and 2 standard error (s.e.) of the mean for 2 weight percent (wt%) SiO₂ bins.

The geochemistry of igneous rocks is known to correlate with tectonic setting¹⁵, arising in part from different styles of mantle melting: water-driven flux melting in subduction zones (volcanic arcs) versus decompression melting in rifts^{16–19}. These contrasting processes form the basis of the classically defined calc-alkaline versus tholeiitic differentiation series, as observed in both major- and trace-element systematics^{16,19–21}. Since the proportion of volcanic versus plutonic magmatism may vary with tectonic setting, we have examined differentiation trends separately for arc and rift samples (based on geographical location for samples predominantly less than 200 Myr old; Methods, Extended Data Figs 2, 3) to prevent such an effect from biasing our geochemical analysis. The results (Figs 1, 2) are shown along with mid-ocean ridge (MOR) differentiation trends as an endmember tholeiitic case for comparison. Samples whose original tectonic environment could not be discerned (grey in Extended Data Fig. 2b) were excluded from plots of arc versus rift geochemistry (Figs 1–3), but included in subsequent differentiation figures.

¹Department of Geosciences, Guyot Hall, Princeton University, Princeton, New Jersey 08544, USA. ²Department of Earth, Planetary, and Space Sciences, University of California, Los Angeles, California 90095-1567, USA. ³Department of Geoscience, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.

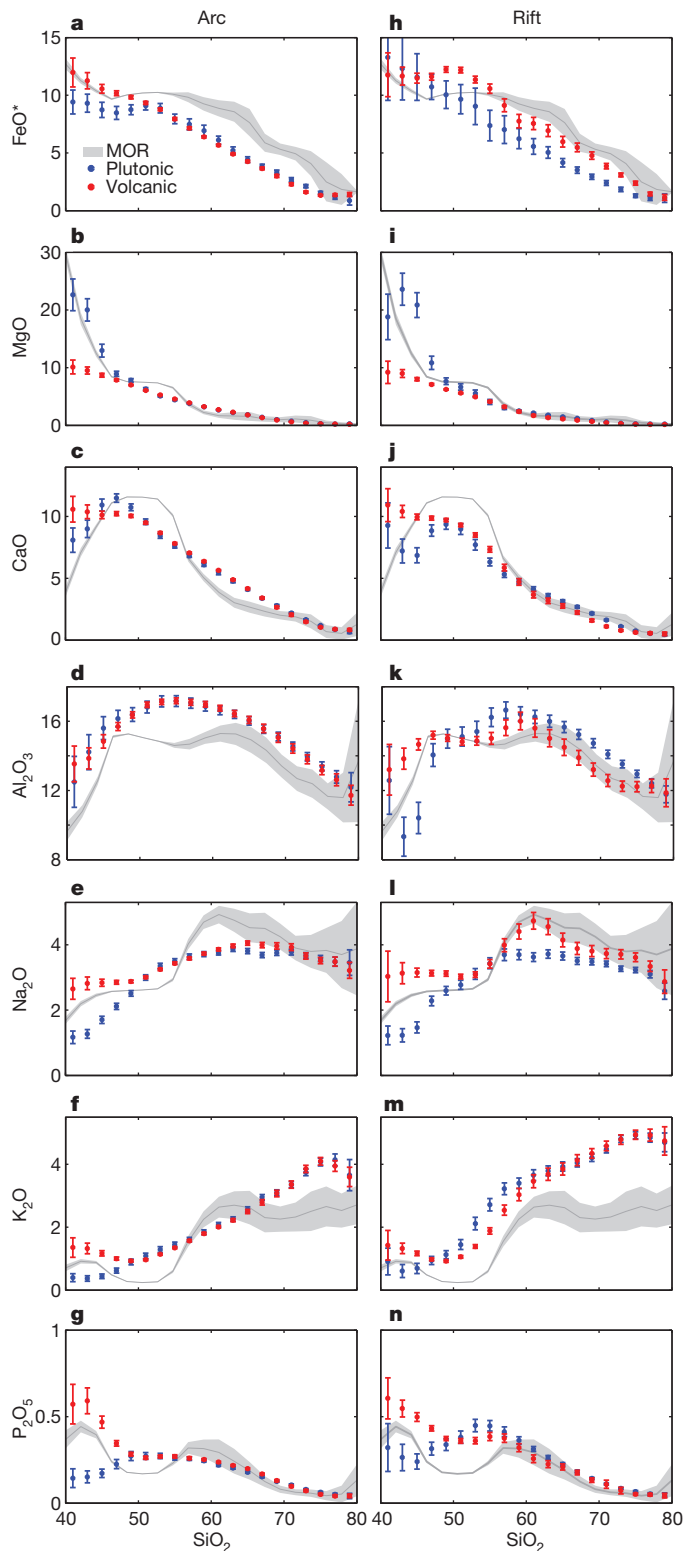


Figure 1 | Major-element Harker diagrams. Shown are mean and 2 s.e. uncertainties of the mean in 2% SiO₂ bins for plutonic (blue) and volcanic (red) samples from arc (a–g; left column) and rift (h–n; right column) tectonic settings. Mean and 2 s.e. confidence intervals for average MORB and felsic MOR differentiates are shown in grey for comparison. Throughout the figures, major elements (given as oxides) are reported as wt%, while trace elements (given as element only) are reported as p.p.m.

Major and trace elements from both tectonic settings follow curvilinear differentiation trends (Figs 1, 2), with rift volcanics falling closer to the tholeiitic MOR data set. Elemental abundances

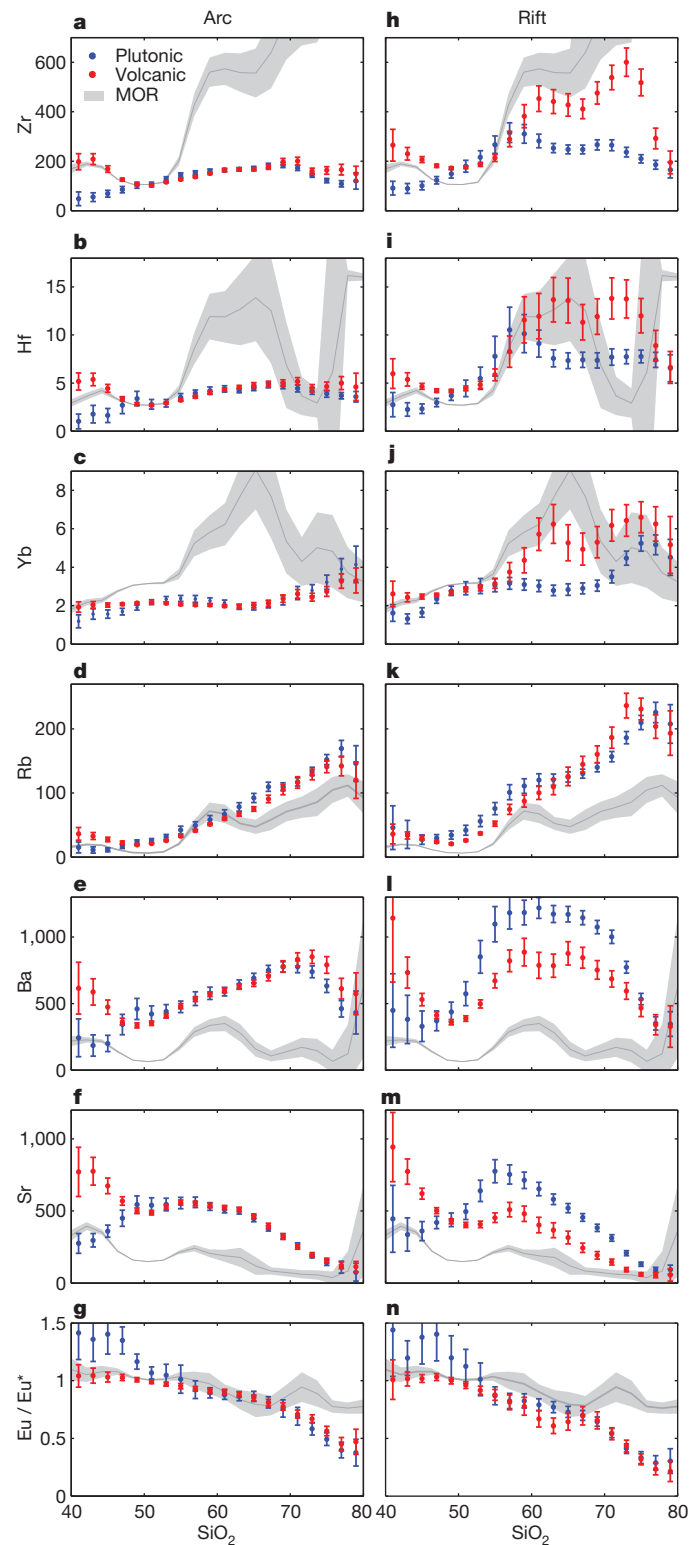


Figure 2 | Trace-element Harker diagrams. As Fig. 1 but for trace elements. Europium anomaly Eu/Eu* represents depletion (Eu/Eu* < 1) or enrichment (Eu/Eu* > 1) of Eu²⁺ relative to adjacent trivalent rare earth elements.

consistently diverge for compositions more mafic than average primitive basalt (that is, <50% SiO₂), with plutonic rocks enriched in MgO and depleted in nominally incompatible major elements (Fig. 1), consistent with the exposure of mafic plutonic cumulates. Enrichments in volcanic TiO₂ and K₂O at low silica may also suggest the influence of highly eruptible peralkaline volcanics with few intrusive equivalents. Above 55% SiO₂, elemental abundances for arc plutonic and volcanic

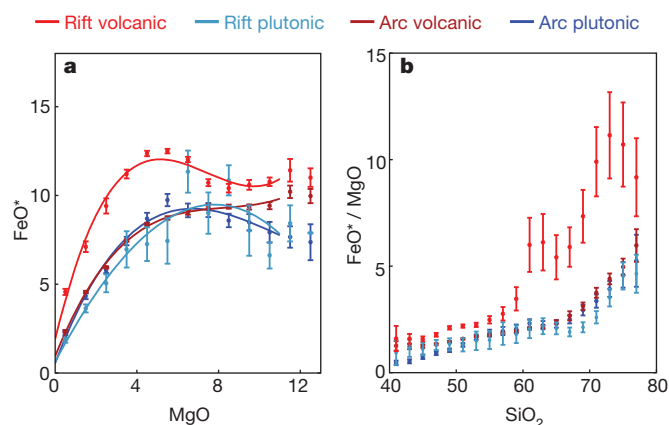


Figure 3 | Fe–Mg systematics of arc and rift volcanic and plutonic rocks. Shown are mean and 2 s.e. uncertainties of FeO^* (total iron normalized as Fe^{2+} oxide) and FeO^*/MgO ratio as functions of MgO and SiO_2 abundance, respectively. **a**, FeO^* binned as a function of MgO in 1 wt% intervals. Increasing FeO^* during differentiation from 8% MgO to 4% MgO indicates a tholeiitic differentiation trend, while decreasing FeO^* indicates a calc-alkaline trend. **b**, FeO^*/MgO binned as a function of SiO_2 in 2 wt% intervals. Greater FeO^*/MgO at a given silica content indicates more tholeiitic differentiation.

rocks follow overlapping trends at the 2 s.e. level, while rift volcanics reveal subtle enrichments in FeO and depletions in CaO and Al_2O_3 relative to rift plutonic samples. More dramatic differences between rift volcanic and plutonic rocks are observed in the trace-element data, with offsets by up to a factor of two across a wide range of SiO_2 . Ba and Sr are significantly enriched in plutonic relative to volcanic rift rocks above 50% SiO_2 (Fig. 2), while conversely, felsic rift plutonic rocks show depletions in Zr, Hf, and the heavy rare earth elements (HREEs). Divergences in rift samples remain when sampling only rocks with known ages <100 Myr (Extended Data Fig. 5), such that plutonic compositions are not biased towards older, basement continental crust. In addition, while rift samples are generally associated with high heat fluxes²² that may permit partial melting of pre-existing crust to produce incompatible-element-enriched melts, average rift differentiation trends in Fig. 1 do not show a strong tendency towards S-type²³ signatures, for example, with lower Al at high silica than in arcs.

As Zr and Hf are nominally incompatible in all major rock-forming minerals, the inflection towards decreasing Zr and Hf with increasing silica reflects saturation of zircon (ZrSiO_4)²⁴. Yttrium and the HREEs are similarly influenced by zircon fractionation, as well as by fractionation of amphibole and garnet^{25,26}. The large divalent cations Ba and Sr, meanwhile, are primarily compatible in feldspar, particularly high-Na plagioclase²⁶. Consequently, decreased feldspar fractionation and increased zircon fractionation in rift plutonic relative to rift volcanic magmas could produce the observed offsets in Zr, Hf, HREEs, and Sr at a given SiO_2 , consistent with a less dramatic but systematic depletion of CaO and Al_2O_3 in rift volcanics compared to plutonics (Fig. 1).

The role of water in magma eruptibility

In one scenario for mineral fractionation between volcanic and plutonic rocks, eruptible felsic melt is extracted from a crystal-rich magma shortly before eruption, leaving behind a lower- SiO_2 , cumulate-like plutonic residue¹¹. This mechanism may be expected to enrich the plutonic residue, as is observed, in Ba and Sr due to feldspar accumulation. However, it is not obvious why such a process would be active in rift environments but completely absent in arcs.

Differences between rift and arc magmas indicate a higher proportion of anhydrous decompression melting in rifts compared to dominantly fluid driven flux melting beneath arcs^{17–19}. As seen in Fig. 1, rift volcanics consistently plot closer to the dry, tholeiitic MOR

differentiation trend than other continental igneous rocks, with increased FeO^* and TiO_2 and decreased CaO and Al_2O_3 at a given silica content^{19,21} (FeO^* is defined in Fig. 3 legend). The tholeiitic character of rift volcanics relative to all other continental samples is illustrated most prominently in Fig. 3, in which rift volcanic rocks show initial FeO enrichment with decreasing MgO that typifies the tholeiitic differentiation trend^{19,21}. In contrast, rift plutonic and arc rocks show lower FeO^* at a given MgO and no FeO enrichment with decreasing MgO. Given the strong correlation between water content and the degree of tholeiitic versus calc-alkaline differentiation^{20,21,27}, this observation is consistent with a systematically more hydrous composition for plutonic rift relative to volcanic rift samples.

Dissolved water in silicate melts has the effect of lowering the liquidus and solidus temperatures such that most silicate minerals become saturated in the magma at lower temperatures^{20,27–30}, with a particularly strong effect for sodic plagioclase^{20,28}. In contrast, experiments reveal little to no effect of water content on the saturation temperature of minerals such as magnetite³¹ and zircon²⁴. Thus, at high water activity, crystallization of common silicates will be suppressed relative to these phases at a given temperature: zircon and magnetite are therefore expected to saturate at lower magma SiO_2 contents during hydrous fractionation. Increased fractional crystallization of magnetite relative to plagioclase and other silicates in hydrous magmas is responsible for FeO depletion in the calc-alkaline trend^{20,21,27,31}. We argue that given the dominant control of zircon on magma Zr and Hf, and combined with decreased partitioning of Ba and Sr in anorthite compared to albite²⁶, calc-alkaline differentiation also explains enrichment of Ba and Sr and depletion of Zr, Hf, and HREEs in the melt phase relative to differentiation at lower water contents.

If Ba and Sr contents are elevated in hydrous magmas owing to suppressed plagioclase stability, the Eu anomaly (Eu/Eu^* ; Fig. 2), typically viewed as an indicator of plagioclase fractionation during differentiation, may be expected to follow a less negative trend for hydrous differentiation. In apparent contradiction to the plagioclase fractionation hypothesis, the data sets show that Eu/Eu^* is identical in all plutonic and volcanic rocks above 52% SiO_2 regardless of tectonic setting. Surprisingly, however, arc Eu/Eu^* is observed to follow equally or even more negative differentiation trends than MOR samples (Fig. 2). The relative lack of negative Eu/Eu^* in dry MOR tholeiites has been observed in other data compilations³², and may suggest that water availability in a differentiating magma has little effect on resulting Eu/Eu^* . Indeed, while plagioclase is often credited as the primary driver of magma Eu/Eu^* , a compilation of published mineral/melt partition coefficients²⁶ illustrates that, for instance, orthoclase may often have a higher $\text{Eu}^{2+}/\text{Eu}^{3+}$ partition coefficient ratio than plagioclase (Extended Data Table 1). Consequently Eu/Eu^* is not directly diagnostic of plagioclase fractionation, and the lack of a more pronounced Eu anomaly in MOR and rift volcanic samples does not rule out increased plagioclase crystallization in these dry lithologies relative to arc and plutonic samples if coupled with complementary H_2O -induced changes in the fractionation of other Eu/Eu^* -active minerals.

Given the high fluid solubility of Ba and Sr^{16,33}, source enrichment by aqueous fluids provides an alternative origin of enhanced Ba and Sr concentrations in rift plutonics. Such metasomatic fluids may be released along with fluid-mobile elements such as Ba, Sr, Pb, and Rb^{16,33} by rift-associated heating of mantle lithosphere or hydrous continental crust. However, enrichment in fluid-soluble elements cannot solely explain Ba and Sr offsets in rift plutonics given the lack of equivalent offsets in other fluid-soluble elements like Rb and Pb (Fig. 2), which are observed to be strongly elevated in known cases of fluid-mobile element enrichment in metasomatic veins and aureoles¹⁶, and in experimental measurements of aqueous fluids produced by dehydration of oceanic crust³³. Melts enriched by such fluids would be driven towards calc-alkaline differentiation trends

by the increased water content, with suppressed plagioclase fractionation and increased zircon fractionation generating a majority of observed volcanic–plutonic offsets. Enrichment by metasomatic fluids may provide an initial water source for hydrous plutonic magmas in rifts, but this cannot yet be conclusively confirmed or rejected with the present data set.

Examination of the hydrous and anhydrous solidi for granitic compositions predicts a plutonic fate for hydrated granitoid magmas. Given the negative slope of the water-saturated solidus, hydrous magmas approach the solidus and may freeze upon decompression due to devolatilization, while decompression of dry magmas moves them away from the solidus towards greater degrees of partial melt (Fig. 4; Supplementary Discussion)^{9,10,29,30}. Consequently, water content may correlate with emplacement depth in settings where magmas with highly variable water contents are observed. In arcs, where magmas are almost ubiquitously hydrated³⁴ and often water-saturated by upper crustal depths for felsic compositions, the probability of eruption may instead be dictated by variables such as magma flux, crustal stress and thermal state. These quantities have already been suggested to control eruption on the basis of geochronological, thermal, and structural considerations^{12,35,36} and arguably must do so in arc settings given identical arc volcanic and plutonic magma compositions. Therefore, while dry tholeiitic magmas have been documented in arc settings¹⁹, our analysis indicates that such melts do not affect average arc differentiation trends.

Generation of felsic continental crust

Despite the critical role of the continental crust on Earth, the relative contribution of two endmember processes responsible for crustal differentiation—fractional crystallization of mantle-derived basaltic magma and partial melting of older crust—remains uncertain^{37–40}. The distribution of silica in crustal rocks, both globally (Fig. 5) and regionally, reveals a pronounced gap in the abundance of intermediate compositions, generally referred to as the ‘Daly Gap’^{37–39,41,42}. This

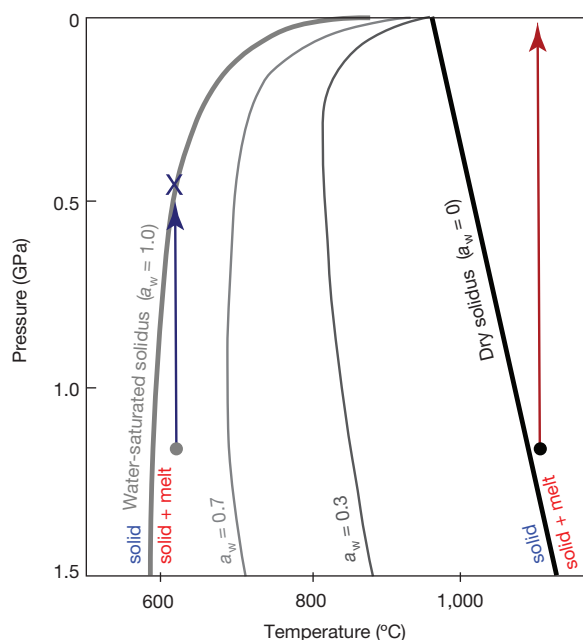


Figure 4 | Water-saturated and water-undersaturated solidi in granitic systems. Wet magmas, which are typically colder than dry magmas due to generation near the solidus of their protolith (Supplementary Discussion), move towards the solidus and may freeze upon decompression^{9,10}, while drier magmas with higher solidus temperatures move away from the solidus and may not crystallize until after eruption. Graph adapted from ref. 30 (GSA), showing granitic solidi for water activities (a_w) ranging from 1 (water-saturated) to 0 (dry).

bimodality initially suggests partial melting as the dominant mechanism of differentiation³⁷: low-degree melting of mantle-derived basalt with $\sim 50\%$ SiO_2 can produce granodioritic and trondhjemitic melts with $\sim 70\%$ SiO_2 or more—corresponding to the two peaks at ~ 50 and $\sim 70\%$ SiO_2 in the observed silica distribution^{27,43}. In this model, intermediate magmas are produced primarily by magma mixing between basaltic and granitic endmembers⁴⁴, as suggested by the observation of crystal-rich andesites⁴². Comparatively, fractional crystallization is less clearly predisposed to produce a bimodal silica distribution. Mechanisms have nonetheless been suggested by which compositional gaps may result from the systematics of crystal fractionation³⁸ or the dynamics of liquid–crystal separation³⁹.

Fractional crystallization and partial melting models may be distinguished in our data set, however, since the two models result in contrasting major- and trace-element differentiation trends at intermediate compositions. Given that intermediate magmas are produced by mixing between granitic and basaltic endmembers in the partial

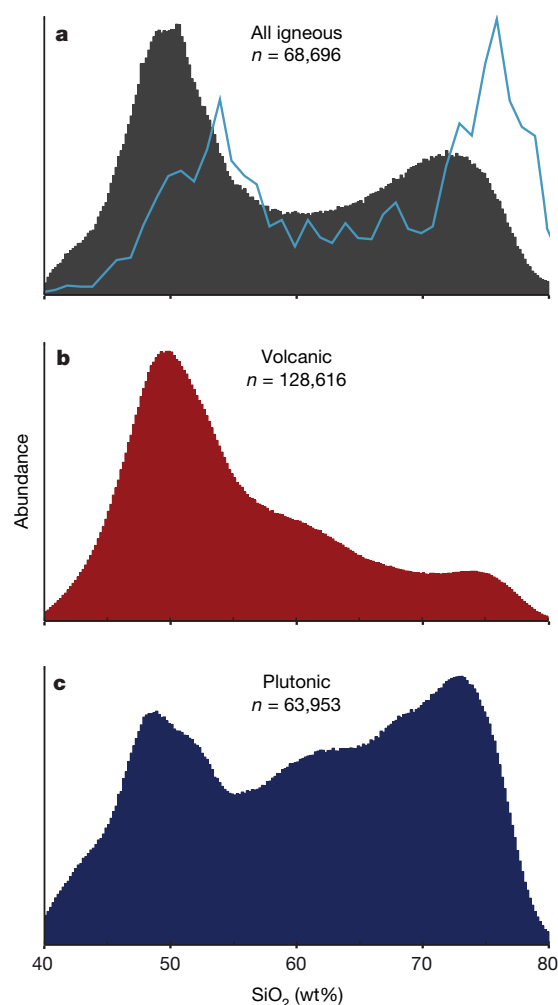


Figure 5 | Histograms of rock abundance as a function of silica content. a–c, Data are shown for all igneous (a), volcanic (b) and plutonic (c) samples. While first observed in oceanic alkaline magma series^{37,38,41}, compositional gaps clearly emerge in a spatiotemporally weighted whole-rock global data set (a) with peak compositions equally spaced but systematically 5 wt% SiO_2 lower than those observed in the melt inclusion histograms of ref. 42 (blue line). The low-silica peak is dominant in volcanic samples (b), while the high-silica peak predominates in plutonic samples³⁰. Since the volcanic record is more influenced by preservation and exposure bias than the plutonic record, the spatiotemporally weighted record of ref. 14 is used in ref. 30 to obtain a better estimate of average time-integrated continental crust, despite the smaller size of this data set.

melting model, differentiation trends in this model must follow straight mixing lines, as physical mixing between two compositional endmembers invariably produces straight lines in element–element space⁴⁵. Fractional crystallization, in contrast, produces curved trends through continuous differentiation^{46–49}.

Consistent with a fractional crystallization model, the observed volcanic and plutonic differentiation trends follow distinctly curved paths, even at intermediate compositions (Figs 1, 2). These curved paths resemble liquid lines of descent for fractional crystallization, overlapping with experimental batch fractional crystallization results (Fig. 6a)^{47–49}. Comparison of observed trends to the differentiation trends of primitive basalt subject to varying degrees of magma mixing suggest a maximum 40% contribution of magma mixing to the production of intermediate magmas relative to idealized continuous fractional crystallization (Fig. 6a inset), decreasing to almost no allowed mixing relative to the experimental batch results. Abundant field and petrological evidence for magma mixing^{44,50} in turn suggests an underlying (pre-mixing) differentiation trend with greater curvature than the observed record, and thus a fractionation mechanism intermediate in efficiency (on average) between batch and continuous fractional crystallization.

One first-order prediction of the fractional crystallization model is the production of abundant crystal cumulates as a by-product of magma differentiation. Such cumulates will vary in mineralogy and composition as a function of the pressure, temperature, and extent of melt extraction⁴⁶, but must be compositionally complementary to the extracted melt, resulting in divergent cumulate and melt compositions⁵¹. As a result, the presence of cumulates in the plutonic record will produce observable compositional offsets between volcanic and plutonic averages for any nonlinear differentiation trends (see, for example, Extended Data Fig. 4). Clear compositional divergence attributable to plutonic cumulates is observed below 50% SiO₂ in both arc and rift settings (for example, MgO, Eu/Eu*). However, compositional offsets between felsic (>62% SiO₂) volcanic and plutonic differentiation trends are absent in arcs (Figs 1, 2) and are

parsimoniously explained in rifts by the petrological effects of water. This lack of a felsic cumulate signature suggests that any cumulates with more than 60% SiO₂ are either small in volume, minimal in degree of melt extraction, or extensively remobilized into the volcanic record—and thus of limited significance for crustal differentiation. Consequently, we propose that the production of even felsic magmas by fractional crystallization results in a predominantly mafic cumulate residue with <62% SiO₂.

To test this prediction, we conducted geochemical modelling in which cumulate compositions, volatile contents, and pressure–temperature (*P–T*) paths representative of average crustal differentiation are estimated by inversion of 1.36 million pMELTS^{52,53} simulations to fit the observed major-element differentiation paths given in Figs 1 and 2 (Methods). As displayed in Fig. 6b–e, felsic magmas with up to 80% SiO₂ are produced in equilibrium with a cumulate of no more than 60% SiO₂. In addition to anticipated outputs of MgO-enriched ultramafic olivine-clinopyroxene-garnet-spinel cumulates, MELTS modelling elucidates the origin of several previously unexplained features of our data set (for example, subtle plutonic K₂O enrichment between 52% and 62% SiO₂ due to orthoclase accumulation; Fig. 6e) while reproducing observed differentiation trends.

The production of intermediate and felsic crust by fractional crystallization is therefore consistent with both the curvature of observed differentiation trends and the existence of mafic cumulates in the plutonic record. However, questions remain regarding crustal silica distributions in a fractional crystallization scenario. As shown in Fig. 5, the compositional gap in crustal igneous rocks is largely driven by the contrasting silica distributions of volcanic and plutonic rocks, with a high-silica peak in plutonic samples and a low-silica peak in volcanic samples. This relationship is the opposite of that predicted by proposed mechanisms of compositional gap formation by mineral–melt separation^{38,39}, and may indicate that the global whole-rock compositional gap is distinct from that observed in some individual volcanic suites^{39,41}. The global Daly Gap as reflected in exposed igneous rocks appears to originate instead from

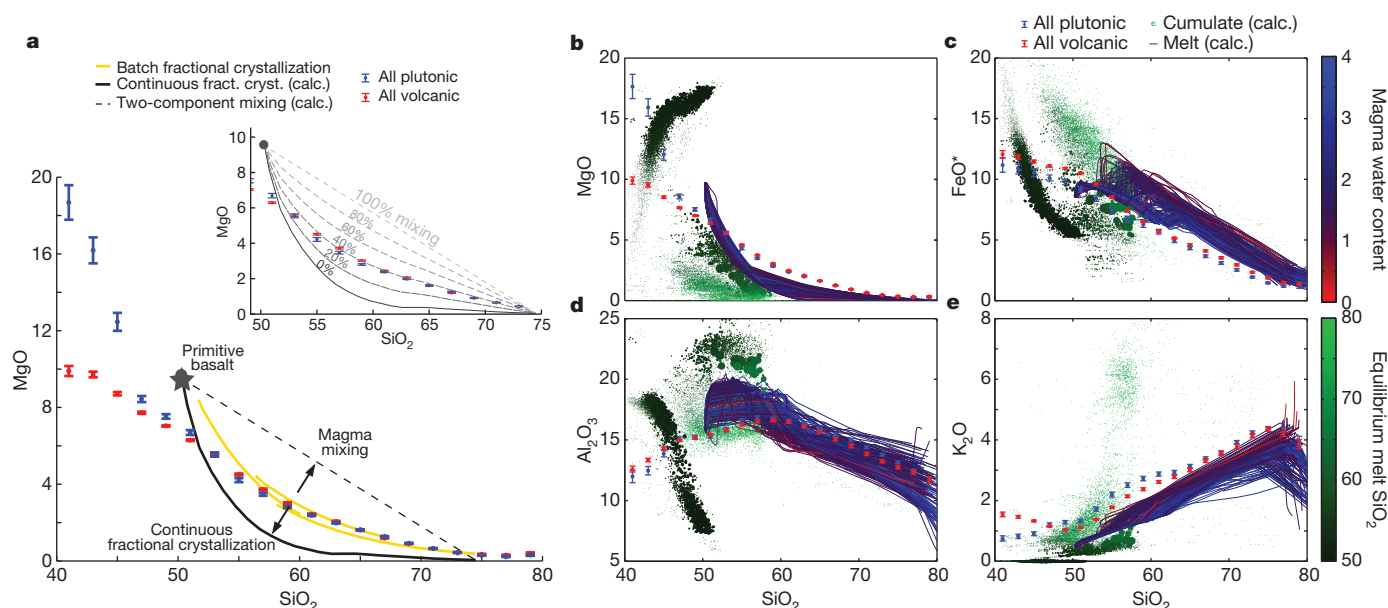


Figure 6 | Comparison of global volcanic and plutonic differentiation to calculated and experimental fractional crystallization trends. Samples are from all tectonic settings (including arcs, rifts, shields, cratons and collisional orogens). **a**, Comparison of volcanic and plutonic MgO differentiation to experimental batch fractionation (yellow, from refs 47–49 as compiled by ref. 42) and calculated continuous fractionation trends (black, average of individual trends in **b** from MELTS). **b–e**, Results of 200 best-fitting MELTS

simulations showing liquid fractionation trends (blue lines) and corresponding cumulate compositions (green points), with point size proportional to the volume of cumulates produced at each simulation step. Cumulate compositions are coloured according to the silica content of the magma with which they are in equilibrium, but are plotted at their own (solid) SiO₂. Minimization is performed relative to volcanic path to avoid potential cumulate influence on melt composition.

superimposing a mantle-derived primitive basaltic distribution with a peak at ~52% SiO₂ onto a differentiated plutonic silica distribution skewed towards high silica.

While mafic cumulates clearly exist in the plutonic record, the higher average silica of plutonic samples is significant in the context of differentiation by fractional crystallization, and suggests that the majority of the mafic residue from crustal differentiation is lost from or otherwise not represented in the plutonic record. This is one expression of a widely-acknowledged problem in crustal mass balance: namely, the production of continental crust with ~61% SiO₂ from basaltic mantle inputs with ~50% SiO₂ (refs 2, 54). This issue is not unique to a fractional crystallization origin of the felsic crust, but can be averted in some crustal melting models by mechanisms where the mafic residue never leaves the mantle, for example, slab melting⁴³. In a fractional crystallization model (and most crustal melting models), however, a mechanism by which cumulates are physically removed from the continental crust is required—for example, lower crustal delamination⁵⁴.

Further implications for the evolution of the crust–mantle system arise from the observed differences in volcanic and plutonic differentiation in rift settings. Because volcanic rocks are more easily eroded and removed from the crust, the tendency towards higher water in rift plutonic rocks should bias the mean crustal composition towards calc-alkaline, thus downplaying the importance of intracontinental magmatism to crustal growth. Meanwhile, the weatherable volcanics are more readily recycled and incorporated into surface biogeochemical cycles—for instance, elevated phosphorus in mafic volcanics (Fig. 1) will necessarily contribute to increased phosphate availability over geological timescales. Conversely, the preferential sequestration of elements enriched in plutonic rocks in the continental crust, such as K, Ba, and Sr (Figs 1, 2), when compounded over billions of years, influences the crust–mantle elemental budget and radiogenic isotope mass balance, and provides additional constraint for quantitative models of crustal growth from the Archaean to the present.

Although our results emphasize the distinction between arc and rift magmatism, the curvature of differentiation trends underscores the importance of fractional crystallization in both environments. Given the exclusively Phanerozoic age of surveyed arc and rift samples, our results suggest that it is possible to produce a bulk andesitic crust by processes of fractional crystallization and subsequent loss of mafic cumulates in modern tectonic settings, consistent with production of the continental crust by tectonic processes similar to those now in operation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 October 2014; accepted 18 May 2015.

- Taylor, S. R. Growth of planetary crusts. *Tectonophysics* **161**, 147–156 (1989).
- Rudnick, R. L. & Fountain, D. Nature and composition of the continental crust: a lower crustal perspective. *Rev. Geophys.* **33**, 267–309 (1995).
- Rosenblatt, P., Pinet, P. C. & Thouvenot, E. Comparative hypsometric analysis of Earth and Venus. *Geophys. Res. Lett.* **21**, 465–468 (1994).
- Walker, J. C. G., Hays, P. B. & Kasting, J. F. A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *J. Geophys. Res.* **86**, 9776–9782 (1981).
- Caldeira, K. Long-term control of atmospheric carbon dioxide; low-temperature seafloor alteration or terrestrial silicate-rock weathering? *Am. J. Sci.* **295**, 1077–1114 (1995).
- Froelich, P. N., Bender, M. L., Luedtke, N. A. & Heath, G. R. The marine phosphorus cycle. *Am. J. Sci.* **282**, 474–511 (1982).
- Hofmann, A. W. Mantle geochemistry: the message from oceanic volcanism. *Nature* **385**, 219–229 (1997).
- Christensen, N. I. & Mooney, W. D. Seismic velocity structure and composition of the continental crust: a global view. *J. Geophys. Res.* **100**, 9761–9788 (1995).
- Carmichael, I. S. E. The andesite aqueduct: perspectives on the evolution of intermediate magmatism in west-central (105–99°W) Mexico. *Contrib. Mineral. Petrol.* **143**, 641–663 (2002).
- Miller, C. F., McDowell, S. M. & Mapes, R. W. Hot and cold granites? Implications of zircon saturation temperatures and preservation of inheritance. *Geology* **31**, 529–532 (2003).
- Bachmann, O., Miller, C. F. & de Silva, S. L. The volcanic–plutonic connection as a stage for understanding crustal magmatism. *J. Volcanol. Geotherm. Res.* **167**, 1–23 (2007).
- Annen, C. From plutons to magma chambers: thermal constraints on the accumulation of eruptible silicic magma in the upper crust. *Earth Planet. Sci. Lett.* **284**, 409–416 (2009).
- EarthChem. Portal Search. (accessed, <http://ecp.iedadata.org/>) (October 2012).
- Keller, C. B. & Schoene, B. Statistical geochemistry reveals disruption in secular lithospheric evolution about 2.5 Gyr ago. *Nature* **485**, 490–493 (2012).
- Pearce, J. A., Harris, N. B. W. & Tindle, A. G. Trace element discrimination diagrams for the tectonic interpretation of granitic rocks. *J. Petrol.* **25**, 956–983 (1984).
- Bebout, G. E. In *Treatise on Geochemistry* Vol. 4 (ed. Rudnick, R. L.) 703–747 (Elsevier, 2014).
- Fyfe, W. S. & McBirney, A. R. Subduction and the structure of andesitic volcanic belts. *Am. J. Sci.* **A 275**, 285–297 (1975).
- Bailey, D. K. The chemical and thermal evolution of rifts. *Tectonophysics* **94**, 585–597 (1983).
- Grove, T. L., Till, C. B. & Krawczynski, M. J. The role of H₂O in subduction zone magmatism. *Annu. Rev. Earth Planet. Sci.* **40**, 413–439 (2012).
- Sisson, T. & Grove, T. L. Experimental investigations of the role of H₂O in calc-alkaline differentiation and subduction zone magmatism. *Contrib. Mineral. Petrol.* **113**, 143–166 (1993).
- Zimmer, M. M. et al. The role of water in generating the calc-alkaline trend: new volatile data for Aleutian magmas and a new tholeiitic index. *J. Petrol.* **51**, 2411–2444 (2010).
- Morgan, P. in *Continental and Oceanic Rifts* (ed. Pálmason, G.) 107–122 (Geodynamics Series Vol. 8, American Geophysical Union, 1982).
- White, A. J. R. & Chappell, B. W. Granitoid types and their distribution in the Lachlan Fold Belt, southeastern Australia. *Geol. Soc. Am. Mem.* **159**, 21–34 (1983).
- Boehnke, P., Watson, E. B., Trail, D., Harrison, T. M. & Schmitt, A. K. Zircon saturation re-revisited. *Chem. Geol.* **351**, 324–334 (2013).
- Schneitzler, C. C. & Philpotts, J. A. Partition coefficients of rare-earth elements between igneous matrix material and rock-forming mineral phenocrysts—II. *Geochim. Cosmochim. Acta* **34**, 331–340 (1970).
- Geochemical. *Earth Reference Model Partition Coefficient Database* (GERM). <http://earthref.org/KDD/> (accessed April 2013).
- Green, T. H. & Ringwood, A. E. Genesis of the calc-alkaline igneous rock suite. *Contrib. Mineral. Petrol.* **18**, 105–162 (1968).
- Yoder, H. S. Jr. Diopside-anorthite-water at five and ten kilobars and its bearing on explosive volcanism. *Yb. Carnegie Inst. Wash.* **64**, 82–89 (1965).
- Burnham, C. W. & Davis, N. F. The role of H₂O in silicate melts; II, Thermodynamic and phase relations in the system NaAlSi₃O₈–H₂O to 10 kilobars, 700 degrees to 1100 degrees C. *Am. J. Sci.* **274**, 902–940 (1974).
- Thompson, A. B. Fertility of crustal rocks during anatexis. *Trans. R. Soc. Edinb.* **87**, 1–10 (1996).
- Botcharnikov, R. E., Almeev, R. R., Koepke, J. & Holtz, F. Phase relations and liquid lines of descent in hydrous ferrobasalt—implications for the Skaergaard intrusion and Columbia River flood basalts. *J. Petrol.* **49**, 1687–1727 (2008).
- Sun, S. & McDonough, W. F. in *Magmatism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313–345 (Geol. Soc. Spec. Publ. Vol. 42, Geological Society, 1989).
- Tatsumi, Y., Hamilton, D. L. & Nesbitt, R. W. Chemical characteristics of fluid phase released from a subducted lithosphere and origin of arc magmas: evidence from high-pressure experiments and natural rocks. *J. Volcanol. Geotherm. Res.* **29**, 293–309 (1986).
- Plank, T., Kelley, K. A., Zimmer, M. M., Hauri, E. H. & Wallace, P. J. Why do mafic arc magmas contain ~4wt% water on average? *Earth Planet. Sci. Lett.* **364**, 168–179 (2013).
- Glazner, A. F., Bartley, J. M., Coleman, D. S., Gray, W. & Taylor, R. Z. Are plutons assembled over millions of years by amalgamation from small magma chambers? *GSA Today* **14**, 4–11 (2004).
- Petford, N., Kerr, R. C. & Lister, J. R. Dike transport of granitoid magmas. *Geology* **21**, 845–848 (1993).
- Chayes, F. Relative abundance of intermediate members of the oceanic basalt-trachyte association. *J. Geophys. Res.* **68**, 1519–1534 (1963).
- Clague, D. A. The oceanic basalt-trachyte association: an explanation of the Daly gap. *J. Geol.* **86**, 739–743 (1978).
- Dufek, J. & Bachmann, O. Quantum magmatism: magmatic compositional gaps generated by melt-crystal dynamics. *Geology* **38**, 687–690 (2010).
- Lee, C.-T. A. & Bachmann, O. How important is the role of crystal fractionation in making intermediate magmas? Insights from Zr and P systematics. *Earth Planet. Sci. Lett.* **393**, 266–274 (2014).
- Daly, R. A. The geology of Ascension Island. *Proc. Am. Acad. Arts Sci.* **60**, 3–80 (1925).
- Reubi, O. & Blundy, J. D. A dearth of intermediate melts at subduction zone volcanoes and the petrogenesis of arc andesites. *Nature* **461**, 1269–1273 (2009).
- Drummond, M. S. & Defant, M. J. A model for trondhjemite-tonalite-dacite genesis and crustal growth via slab melting: Archean to modern comparisons. *J. Geophys. Res.* **95**, 21503–21521 (1990).
- Kent, A. J. R., Darr, C., Koleszar, A. M., Salisbury, M. J. & Cooper, K. M. Preferential eruption of andesitic magmas through recharge filtering. *Nature Geosci.* **3**, 631–636 (2010).
- Langmuir, C. H., Vocke, R. D. Jr, Hanson, G. N. & Hart, S. R. A general mixing equation with applications to Icelandic basalts. *Earth Planet. Sci. Lett.* **37**, 380–392 (1978).
- Bowen, N. L. *The Evolution of the Igneous Rocks* (Oxford Univ. Press, 1928).
- Grove, T. L., Donnelly-Nolan, J. M. & Housh, T. Magmatic processes that generated the rhyolite of Glass Mountain, Medicine Lake volcano, N. California. *Contrib. Mineral. Petrol.* **127**, 205–223 (1997).

48. Müntener, O., Kelemen, P. B. & Grove, T. L. The role of H₂O during crystallization of primitive arc magmas under uppermost mantle conditions and genesis of igneous pyroxenites: an experimental study. *Contrib. Mineral. Petrol.* **141**, 643–658 (2001).
49. Alonso-Perez, R., Müntener, O. & Ulmer, P. Igneous garnet and amphibole fractionation in the roots of island arcs: experimental constraints on andesitic liquids. *Contrib. Mineral. Petrol.* **157**, 541–558 (2009).
50. Eichelberger, J. C., Chertkoff, D. G., Dreher, S. T. & Nye, C. J. Magmas in collision: rethinking chemical zonation in silicic magmas. *Geology* **28**, 603–606 (2000).
51. Gelman, S. E., Deering, C. D., Bachmann, O., Huber, C. & Gutierrez, F. J. Identifying the crystal graveyards remaining after large silicic eruptions. *Earth Planet. Sci. Lett.* **403**, 299–306 (2014).
52. Ghiorso, M. S., Hirschmann, M. M., Reiners, P. W. & Kress, V. C. III. The pMELTS: a revision of MELTS for improved calculation of phase relations and major element partitioning related to partial melting of the mantle to 3 GPa. *Geochem. Geophys. Geosyst.* **3**, 1–36 (2002).
53. Smith, P. M. & Asimow, P. D. Adibat_1ph: a new public front-end to the MELTS, pMELTS, and pHMELTS models. *Geochem. Geophys. Geosyst.* **6**, 1–8 (2005).
54. Kay, R. W. & Kay, S. M. Creation and destruction of lower continental crust. *Geol. Rundsch.* **80**, 259–278 (1991).

Supplementary Information is available in the online version of the paper.

Acknowledgements C.B.K. was supported by a Computational Science Graduate Fellowship through US Department of Energy Office of Science grant DE-FG02-97ER25308. Computational resources were provided by the National Energy Research Scientific Computing Center under DOE Office of Science Contract No. DE-AC02-05CH11231, and by the Princeton Institute for Computational Science and Engineering. The authors are grateful to P. M. Antoshechkina for providing recompiled alphaMELTS executables and J. D. Gronewold for initial contributions.

Author Contributions C.B.K. conducted the Monte Carlo calculations and geochemical simulations. All authors participated in interpretation of the results and preparation of the manuscript.

Author Information Computational source code under a GNU GPL v2.0 open-source license, along with all source data, is freely available at <https://github.com/PrincetonUniversity/VolcanicPlutonic> and as Supplementary Information. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.B.K. (cbkeller@princeton.edu).

METHODS

In order to elucidate the geochemical relationship between volcanic and plutonic rocks, we have prepared and analysed a data set of major- and trace-element compositional data for 171,690 volcanic and 122,751 plutonic whole-rock samples together with location and lithological information (Supplementary Data 1). Data were obtained from the freely-accessible EarthChem Portal, including samples from NavDat, Georoc PetDB, and the US Geological Survey¹³. Oceanic samples were extracted to calculate tholeiitic MOR differentiation trends, while continental samples were separated into volcanic and plutonic data sets based on reported lithology. Volcanic and plutonic samples were further classified by cross-correlating sample location with present-day tectonic environment, using a geospatially referenced tectonic setting map modified from the USGS Global Crustal Database⁵⁵ (Extended Data Fig. 2).

As illustrated in Extended Data Fig. 2, only samples from active oceanic and continental arcs are included in the 'arc' category, while samples from active rifts along with some well-established failed rifts (aulacogens) and plume-associated continental large igneous provinces were included in the 'rift' category. Consequently, rift samples represent those whose parental basaltic magma was produced primarily by decompression melting of the mantle, while arc samples are characterized by the contributions of flux melting in the arc mantle wedge. Continental samples whose original tectonic environment could not be discerned (grey area in Extended Data Fig. 2) were excluded from plots of arc versus rift geochemistry, but included in subsequent figures.

In contrast to the data set of Keller and Schoene¹⁴, a substantial proportion (~40%) of the whole-rock samples in our larger volcanic-plutonic data set are not associated with age constraints. Tectonic setting assignments based on present-day location are, naturally, only valid for samples produced in their current setting, limiting the accuracy of assignments for samples older than a few hundred million years. If numerous, mis-assigned samples would probably blur the distinction between arc and rift subsets. However, based on the distribution of known sample dates (Extended Data Fig. 3), the proportion of older 'inherited' samples (>300 Ma) in modern arc and rift settings is minor in our data set (<10%). This dominantly Phanerozoic age distribution has the additional benefit of increasing the probability that the process of tectonics and crust formation has not changed within the period of our analysis. Accordingly, differentiation diagrams including only samples with known ages in the last 100 Myr (Extended Data Fig. 5) display the same trends and offsets as those including all available data, only with somewhat larger standard errors of the mean due to smaller sample size.

Owing to the large size of the data set, spurious data was filtered both manually and algorithmically (Supplementary Data 4). Although outlier removal may introduce bias by rejecting valid but exceptional data, the presence of extreme outliers (for example, several orders of magnitude farther from the median than the 25% CI and probably not representing real bulk rock compositions) necessitated a method of outlier rejection to produce optimally accurate distributions of crustal composition. Consequently, conservative outlier rejection bounds were set manually in severe cases. Data representing physically impossible conditions (for example, compositional normalization >100%) was not considered meaningful, and thus deleted categorically.

Data were then subjected to a modified version of the statistical approach of Keller and Schoene¹⁴. Since geochronological constraints are not available for many samples in the volcanic-plutonic data set, resampling weights were based only on sample location, rather than location and age as in Keller and Schoene¹⁴, to obtain an optimally uniform posterior spatial distribution (Extended Data Fig. 1). Although elemental distributions in the data set are often not Gaussian (Extended Data Figs 6, 7), the arithmetic mean of a set of sample compositions nonetheless has unique physical significance, representing the bulk composition that would be produced by physically homogenizing that sample set. Further, in contrast to the distribution of the elemental data, the distribution of the mean for each element is nearly Gaussian, due to the effects of the central limit theorem. Consequently, Figs 1, 2 show the arithmetic mean of binned volcanic and plutonic samples, along with the two-sigma standard error of the mean; equivalent plots based instead on the median are shown in Extended Data Fig. 6, with congruent results.

As the single dominant oxide component of most igneous rocks, wt% SiO₂ is simpler to interpret and easier to correlate with observed rock compositions than most alternative differentiation indicators. Nonetheless, similar volcanic-plutonic trends to those in the Harker diagrams (Figs 1, 2) may be observed as well with wt% MgO as the independent variable (Extended Data Fig. 8), with decreasing MgO corresponding to increasing differentiation.

MELTS simulations. To test our predictions regarding the influence of fractional crystallization on average differentiation trends, the characteristics of cumulates in the plutonic record, and the effect of water on differentiation, we

have conducted trace- and major-element geochemical modelling using high-performance computational resources at the DOE's National Energy Resource Supercomputing Center and the Princeton Institute for Computational Science and Engineering. The massive parameter space and the nonlinear, discontinuous equations of igneous differentiation processes such as fractional crystallization lead to highly nonunique solutions which cannot be obtained analytically except in the simplest cases. Consequently, we have approached the problem through a brute-force Monte Carlo approach, minimizing 1.36 million MELTS fractionation paths with differing starting compositions and *P-T* paths to fit the observed crustal differentiation trends. These simulations were conducted using the alphaMELTS⁵³ v1.4.1 command-line version of the MELTS and pMELTS thermodynamic modelling software⁵² along with parallel scripting codes (Supplementary Data 4, Supplementary Data 2). As each forward MELTS simulation is independent, the computation is inherently highly parallel, with minimal communication requirements between processes. However, due to the design of MELTS and alphaMELTS, numerous small configuration, input, and output files must be written for each simulation, such that networked filesystem performance may be a limiting factor for scalability. The average primitive oceanic basalt of Kelemen⁵⁶ was used as a starting composition, with varying initial H₂O, CO₂ contents and *P-T* conditions. Initial H₂O and CO₂ contents were drawn from uniform distributions between 0–4% by weight for H₂O and between 0–1% for CO₂. Initial oxygen fugacity was equilibrated at the fayalite–magnetite–quartz buffer (FMQ) but subsequently allowed to evolve in response to mineral fractionation.

Pressure ranges of differentiation were set following the assumption that basaltic magmas must be generated in the mantle (conservatively, at depths greater than ~33 km) and solidify in the crust. Maximum pressure is limited in a typical arc setting by the depth of the subducted slab (~100 km) but was constrained slightly shallower in practice so as not to greatly exceed the calibration range of pMELTS. Consequently, initial and final pressures were drawn from uniform random distributions between 1–2.5 and 0–1 GPa, respectively, with initial temperature set by the primitive basalt liquidus and a final temperature of 700 °C. Pseudorandom *P-T* paths were then generated by selecting five points from a uniform distribution for both pressure and temperature between the previously generated minima and maxima for each simulation, subject to the constraint that subsequent points must be non-increasing in both pressure and temperature. Each *P-T* path was then constructed by linearly interpolating between these seven points. The production simulations were then run using alphaMELTS v1.2 with a fractionation threshold of 0.005 and a 10 °C cooling step size. To explore the effect of oxygen fugacity, an equivalent set of simulations was conducted with an initial oxygen fugacity one log unit above the FMQ buffer (FMQ + 1). The initial volatile compositions, average volatile compositions, and *P-T* paths of the best-fitting 350 simulations out of 1.36 × 10⁶ simulations for initial oxygen fugacities of FMQ and FMQ + 1 are shown in Extended Data Fig. 8, along with a comparison of best-fitting major-element paths for these two different initial redox states. Results elsewhere throughout the text show the FMQ version.

Given the large compositional and physical parameter space sampled, some simulations will inevitably fall in regions where MELTS is poorly calibrated. In addition, saturation of hydrous mafic phases (largely amphibole) is probably underestimated in general, while at high silica MELTS may underestimate the saturation of quartz and overestimate that of potassium feldspar; rhyolite-MELTS, which addresses the latter two problems⁵⁷ (not available in an alphaMELTS scripted version at time of our simulations) may result in different inversion results. Consequently, our simulations should be reproduced with multiple versions of MELTS before assigning particular significance to second-order results such as the inverted optimal *P-T* paths or water contents shown in Extended Data Fig. 8. Nonetheless, the MELTS results do unequivocally demonstrate that fractionation of the calculated cumulate compositions can accurately reproduce the observed average major-element differentiation trends.

Trace-element calculations. In order to additionally simulate trace-element fractionation, we calculated representative average mineral/melt partitioning using partition coefficients from the GERM partition coefficient database²⁶ (Supplementary Data 3). Owing to the strong compositional dependence of most trace-element partition coefficients (mineral/melt partition coefficients tend to increase with increasing magma silicate network strength), partition coefficients were binned and interpolated as a function of magma SiO₂ content, propagating errors through unweighted Monte Carlo bootstrap resampling (Supplementary Data 4).

Trace-element differentiation paths were calculated for ideal fractional crystallization using mineral phase proportions from the best-fit MELTS simulations. As seen in Extended Data Fig. 9, calculated fractionation paths conform to the expectation of increasing Sr, and to a lesser degree Ba with increasing magma water content. As seen in Extended Data Table 1, the compiled partition

coefficient data reveals high K_d Eu/Eu* in orthoclase as well as plagioclase feldspar, resulting in no clear correlation of magma water content with Eu anomaly during differentiation (Extended Data Fig. 9).

Solidus temperature of hydrous and dry magmas. While the opposing slopes of the hydrous versus dry solidi are unambiguously determined from experimental data for a wide range of silicate systems^{30,58}, connecting these solidi to the preferential stalling of hydrous magmas upon decompression also requires lower average temperatures for hydrous magmas. The assumption of colder temperatures in hydrous magmas is somewhat justified since these magmas inevitably derive from systems with depressed solidi and liquidus. As an example, let us consider an arbitrary silicate protolith to which a finite quantity of heat is added, sufficient to produce an extractable proportion of silicate magma over a geologically relevant temporal and spatial scale. Two factors suggest that, in this scenario, final magma temperature will be linked to the solidus and liquidus temperatures of the protolith.

First, for processes within the solid earth, rates of heat transport are competitive with those of heat addition. Once the solidus is reached in one region of the protolith, heat can diffuse and advect to adjacent regions of unmelted protolith, until the available heat is consumed as latent heat of fusion. This is demonstrated by the extreme rarity of terrestrial supersolidus magmas, which are thought to be produced only in cases of extremely rapid heat addition such as large bolide impacts⁵⁹.

Second, the molar heat capacity C_p (in $\text{J mol}^{-1} \text{K}^{-1}$) of most silicates is much lower (by two to three orders of magnitude) than their molar latent heat of fusion ΔH_f (in J mol^{-1})⁶⁰. In this sense, a difference in solidus temperature is equivalent in terms of heat consumed to a comparatively small change in the extent of protolith melting. As long as the liquidus is not exceeded, then, the final magma temperature will be a consistent offset above the solidus over a range of solidus temperatures.

In other words, once the solidus is reached, the remaining available heat is primarily used to produce more magma, not raise magma temperature. We may then infer that the addition of heat to a hydrous (low- T solidus) versus dry (high- T solidus) protolith will result in a relatively similar extent of melting—and thus a similar position relative to the appropriate solidus—but with higher temperature for the dry magma due to its higher solidus temperature.

The same reasoning holds for the addition of water to an existing magma: the addition of water will promote further melting (of entrained crystals, if any, or of the surrounding wall rock), decreasing magma temperature as the melting reaction consumes the available excess enthalpy.

Causality of the plutonic water correlation. If our observed correlation between plutonic emplacement and geochemical signatures of higher water content is not fortuitous, there are at least two plausible causal mechanisms we must consider, with opposite directions of causality: that plutonic magmas are hydrous because they are not degassed near the surface, or that their higher water content forces them to stall at depth.

Given that the solubility of water in silicate magma increases with increasing pressure⁶¹, there is an inherent bias towards higher water content in magmas differentiated and emplaced at higher pressure. However, our geochemical evidence for higher water in plutonic magmas requires increased water content during the period of differentiation, rather than that of final emplacement. The apparent dearth of cumulates in the upper crust (Figs 1, 2) may indicate that proportionately little differentiation occurs in shallow subvolcanic magma chambers, partially decoupling depth of emplacement from depth of differentiation. If, despite these considerations, greater emplacement depth does contribute significantly to higher water content in plutonic magmas during differentiation, this effect and the associated geochemical signatures of higher water content in plutonic magmas during differentiation should be observed equally in both arc and rift environments, which is not what our data show. In contrast, higher water content leading to an increased likelihood of plutonic emplacement invokes no such contradiction if arc magmas are (as expected) uniformly hydrous at crustal depths. Such considerations, combined with a thermodynamic expectation for the stalling of hydrous magmas as discussed in the section on solidus temperature lead us to favour the latter direction (plutonic because hydrous) of causality.

Eu anomalies. Although negative Eu anomalies in magmas are typically attributed to plagioclase fractionation^{62,63}, plagioclase is not the only mineral with substantial potential effect on Eu/Eu*. Instead, as observed in Extended Data Table 1, minerals such as orthoclase, sphene, apatite, allanite, clinopyroxene, and garnet may also substantially both the Eu budget and Eu/Eu* ratio²⁶ (Supplementary Data 3). One potential explanation for the apparent lack of a systematic difference in Eu anomaly between dry and hydrous magmas hinges on the relative abundance of orthoclase and plagioclase feldspar. While water suppresses plagioclase stability (other factors held constant), it also increases the relative proportion—and possibly the absolute proportion—of orthoclase feldspar. Given the substantially greater K_d Eu/Eu* and K_d Eu of orthoclase relative to anorthite (Extended Data Table 1), the increased partitioning of Eu

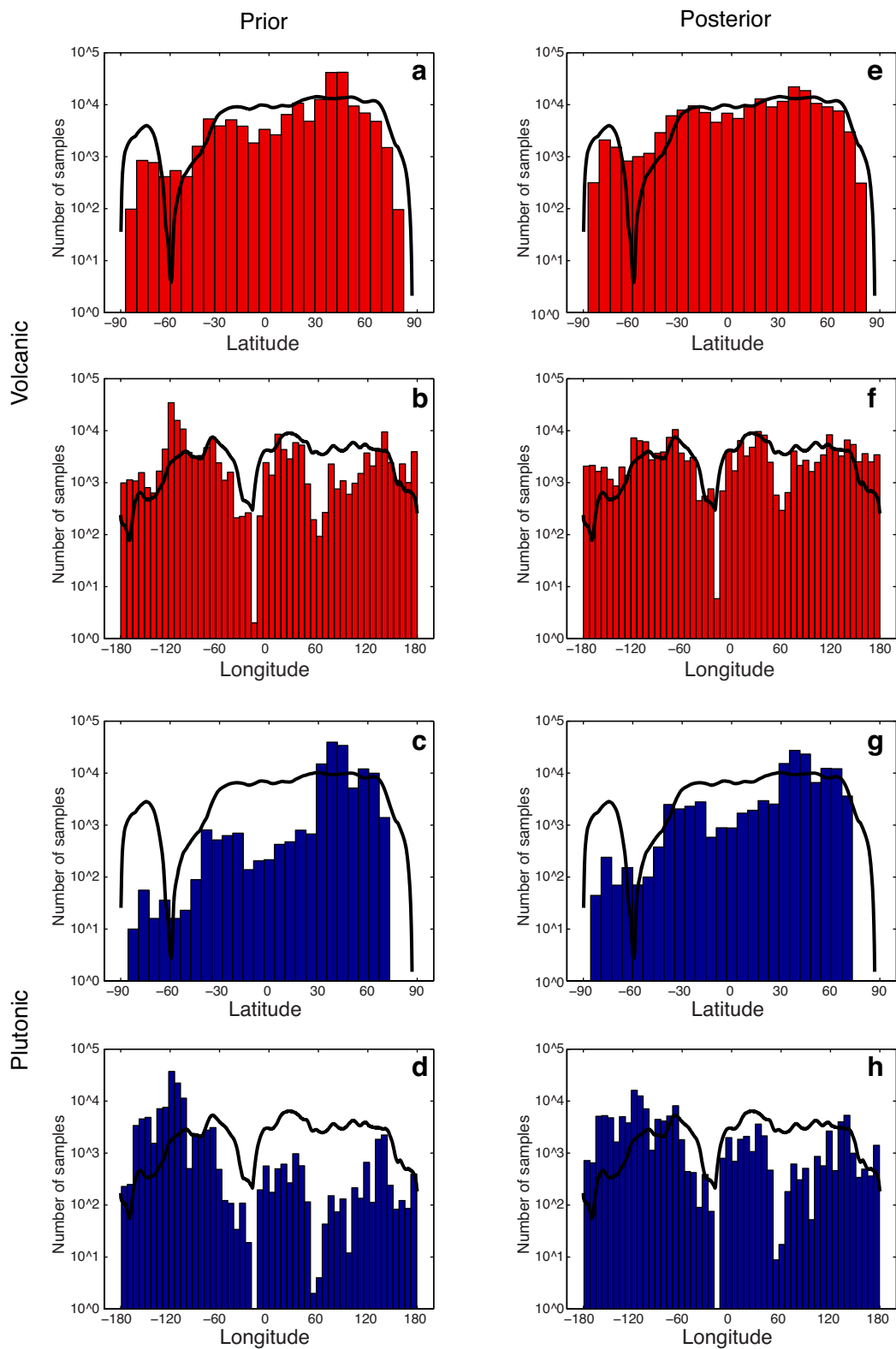
into orthoclase may offset the decrease in Eu fractionation by plagioclase, resulting in an equally negative Eu anomaly of the magma even in the case of decreased plagioclase fractionation. Such an explanation is not without caveats: for instance, considering only the feldspars, this mechanism would predict that any water-related Ba offset produced by plagioclase suppression would be similarly negated by orthoclase, which is not observed. However, the lack of any difference in Eu anomaly between arc and rift (including MOR) samples for a given silica range demonstrates that, whatever the its cumulative effects on mineral stability, water activity evidently does not substantially alter magma Eu anomalies (Fig. 2).

Melt extraction from crystal-rich magmas. One potentially significant mechanism for the production of compositional offsets between volcanic and plutonic compositions is the extraction of mobile (and potentially eruptible) melt from a magma, leaving a cumulate plutonic residue^{11,51}. At low silica, plutonic cumulates would be expected to be enriched in minerals such as olivine and pyroxene relative to non-cumulate volcanic magmas, resulting in enrichments of compatible elements such as Mg and Ni. Meanwhile, accumulation of plagioclase and pyroxene, as in (similarly mafic) anorthosites and cumulate gabbros would be expected to result in positive plutonic Eu anomalies. Evidence of such mafic cumulate processes is evident in the geological record^{64–66} as well as from our analysis of geochemistry in both arc and rift environments (Figs 1, 2).

At higher silica, a similar process of melt extraction would be expected to leave plutonic cumulates enriched in feldspar along with potentially lesser proportions of amphibole, biotite, and quartz. This feldspar accumulation would enrich the felsic plutonic residue in Ba and Sr⁵¹, as is observed in rift settings. However, felsic plutonic cumulates resulting from shallow melt extraction would not readily account for the occurrence of plutonic Ba and Sr enrichments only in rift (and not arc) settings, nor the additional volcanic enrichments of Zr and Hf in rift settings. While plagioclase saturation may be suppressed in arcs relative to rifts due to higher overall water content, plagioclase remains a dominant rock-forming mineral in arcs. Accordingly, suppression of plagioclase in arc settings is evidently not so dramatic as to prevent or even markedly reduce the formation of anorthosite or leucogabbro-like mafic cumulates with positive Eu anomalies in arcs (Fig. 2), and would not be expected to prevent or reduce the formation of feldspathic-cumulates at higher silica when magma temperatures have fallen farther into the feldspar stability range. As such, melt extraction in should be expected to result in Ba and Sr enrichment of the residue in arc settings as much as in rifts. Consequently, although shallow volcanic melt extraction leaving a plutonic residue has been argued to occur in some scenarios⁶⁷ (and our results should not be taken to indicate otherwise), it does not appear to be a volumetrically significant mechanism for the production and differentiation of felsic crust.

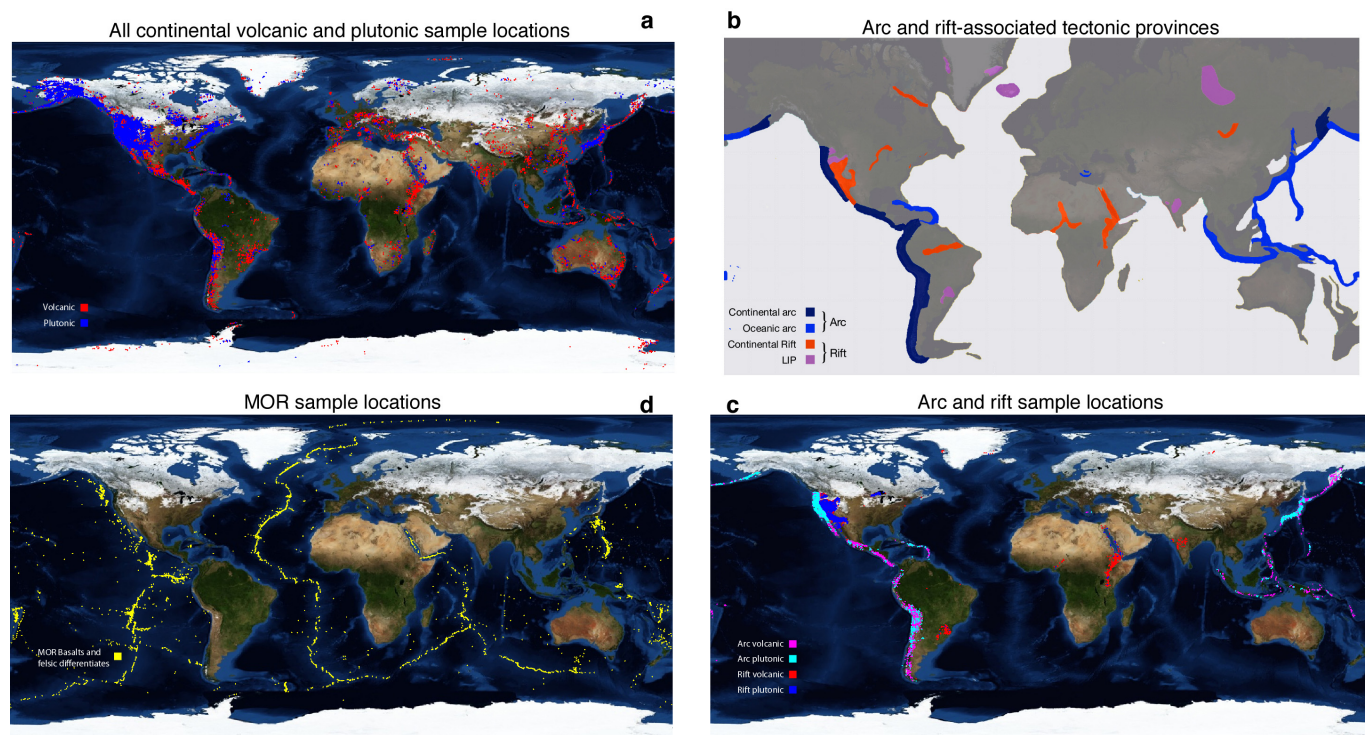
Data and code availability. The full data sets for volcanic and plutonic whole rock composition, partition coefficients, and MELTS simulation output along with all associated source code under a GNU GPL v2.0 open-source license are provided as Supplementary Data, and are freely available on GitHub at <https://github.com/PrincetonUniversity/VolcanicPlutonic>.

55. USGS. Global Crustal Database (US Geological Survey, <http://earthquake.usgs.gov/data/crust/maps.php> (1997).
56. Kelemen, P. B., Hanghøj, K. & Greene, A. R. In *Treatise on Geochemistry* Vol. 4 (ed. Rudnick, R. L.) 749–806 (Elsevier, 2014).
57. Gualda, G. A. R., Ghiorsio, M. S., Lemons, R. V. & Carley, T. L. Rhyolite-MELTS: a modified calibration of MELTS optimized for silica-rich, fluid-bearing magmatic systems. *J. Petrol.* **53**, 875–890 (2012).
58. Green, D. H. Experimental melting studies on a model upper mantle composition at high pressure under water-saturated and water-undersaturated conditions. *Earth Planet. Sci. Lett.* **19**, 37–53 (1973).
59. Zieg, M. J. & Marsh, B. D. The Sudbury Igneous Complex: viscous emulsion differentiation of a superheated impact melt sheet. *Geol. Soc. Am. Bull.* **117**, 1427–1450 (2005).
60. Holland, T. & Powell, R. An improved and extended internally consistent thermodynamic dataset for phases of petrological interest, involving a new equation of state for solids. *J. Metamorph. Geol.* **29**, 333–383 (2011).
61. Hamilton, D. L., Burnham, C. W. & Osborn, E. F. The solubility of water and effects of oxygen fugacity and water content on crystallization in mafic magmas. *J. Petrol.* **5**, 21–39 (1964).
62. Philpotts, J. A. & Schnetzler, C. C. Europium anomalies and the genesis of basalt. *Chem. Geol.* **3**, 5–13 (1968).
63. Weill, D. F. & Drake, M. J. Europium anomaly in plagioclase feldspar: experimental results and semiquantitative model. *Science* **180**, 1059–1060 (1973).
64. Hunter, R. H. & Sparks, R. S. J. The differentiation of the Skaergaard Intrusion. *Contrib. Mineral. Petrol.* **95**, 451–461 (1987).
65. Ashwal, L. D. *Anorthosites* (Springer, 1993).
66. Jagoutz, O. & Schmidt, M. W. The formation and bulk composition of modern juvenile continental crust: the Kohistan arc. *Chem. Geol.* **298–299**, 79–96 (2012).
67. Bachmann, O. On the origin of crystal-poor rhyolites: extracted from batholithic crystal mushes. *J. Petrol.* **45**, 1565–1582 (2004).



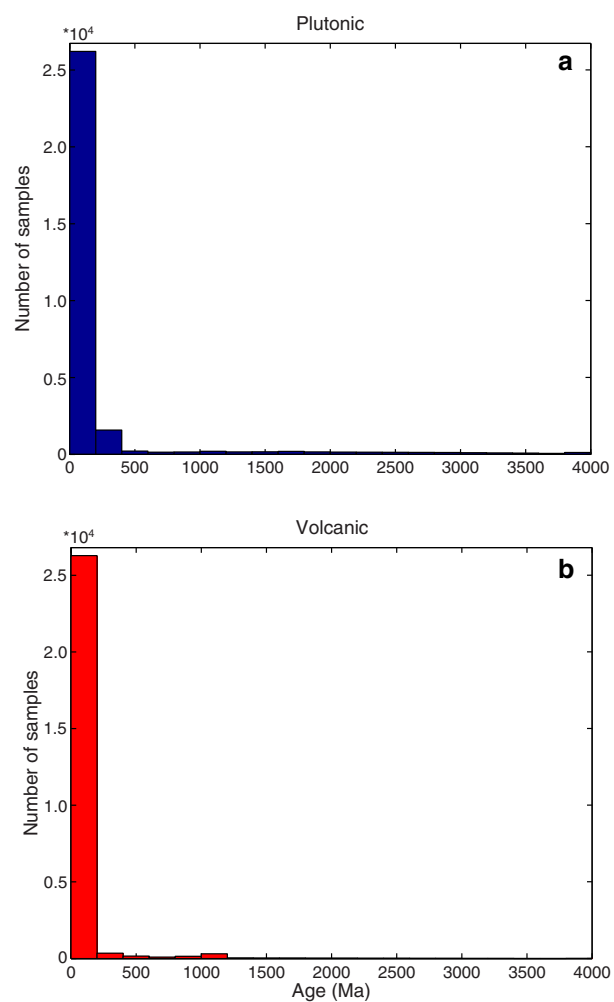
Extended Data Figure 1 | Prior and posterior sample distributions for both volcanic and plutonic samples. Left column, prior distributions (original data, a–d); right column, posterior distributions (that is, after Monte-Carlo analysis, e–h); note log scale on vertical axis. Red, volcanic samples; blue, plutonic

samples; ideal distributions given uniform sampling of the continents (black line) are shown for comparison. Since bootstrap resampling probabilities are weighted to reject no more than 90% of samples, improvement in posterior distribution is limited to roughly one log unit.

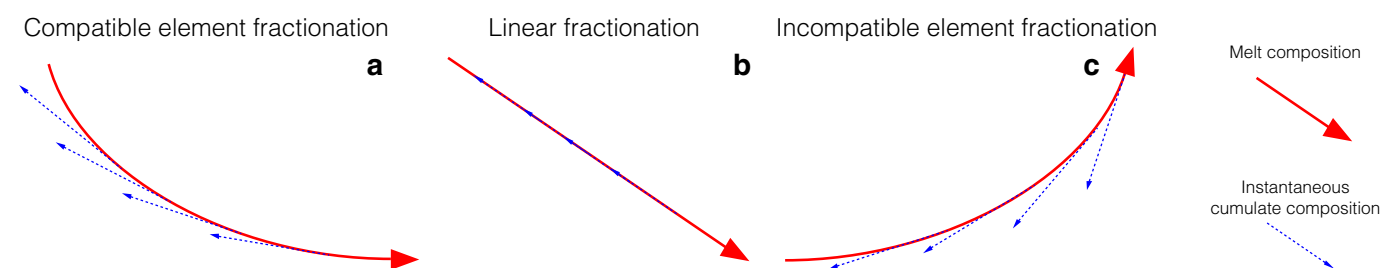


Extended Data Figure 2 | Sample locations and tectonic setting assignments. **a**, All volcanic and plutonic sample locations ($n = 294,347$). **b**, Tectonic setting map modified from the USGS Global Crustal Database (adapted from ref. 55, USGS). **c**, Assigned arc and rift volcanic and plutonic sample locations ($n = 141,561$). **d**, Oceanic (MOR) sample locations

($n = 48,605$). Plutonic samples are plotted above and may obscure full range of volcanic samples. Samples not originating in explicit arc or rift provinces (grey area in **b**) were excluded from setting-specific differentiation trends, but included in the general volcanic and plutonic differentiation trends in subsequent figures.

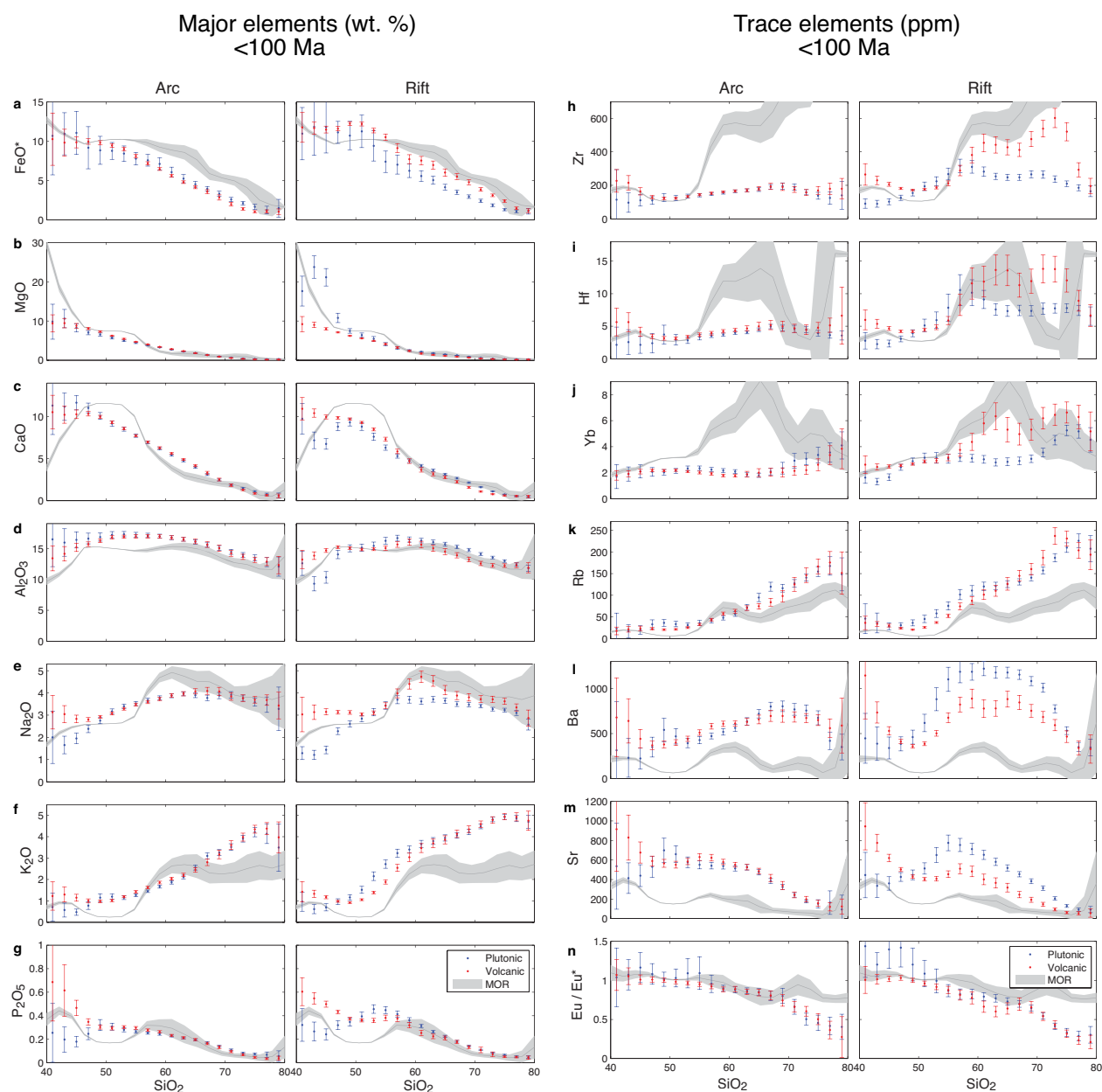


Extended Data Figure 3 | Age distributions for plutonic and volcanic samples with reported age data. a, Plutonic (blue); b, volcanic (red). Both distributions are dominated by Phanerozoic (<540 Ma) samples, with a slightly greater proportion of pre-200 Ma samples in the plutonic distribution.



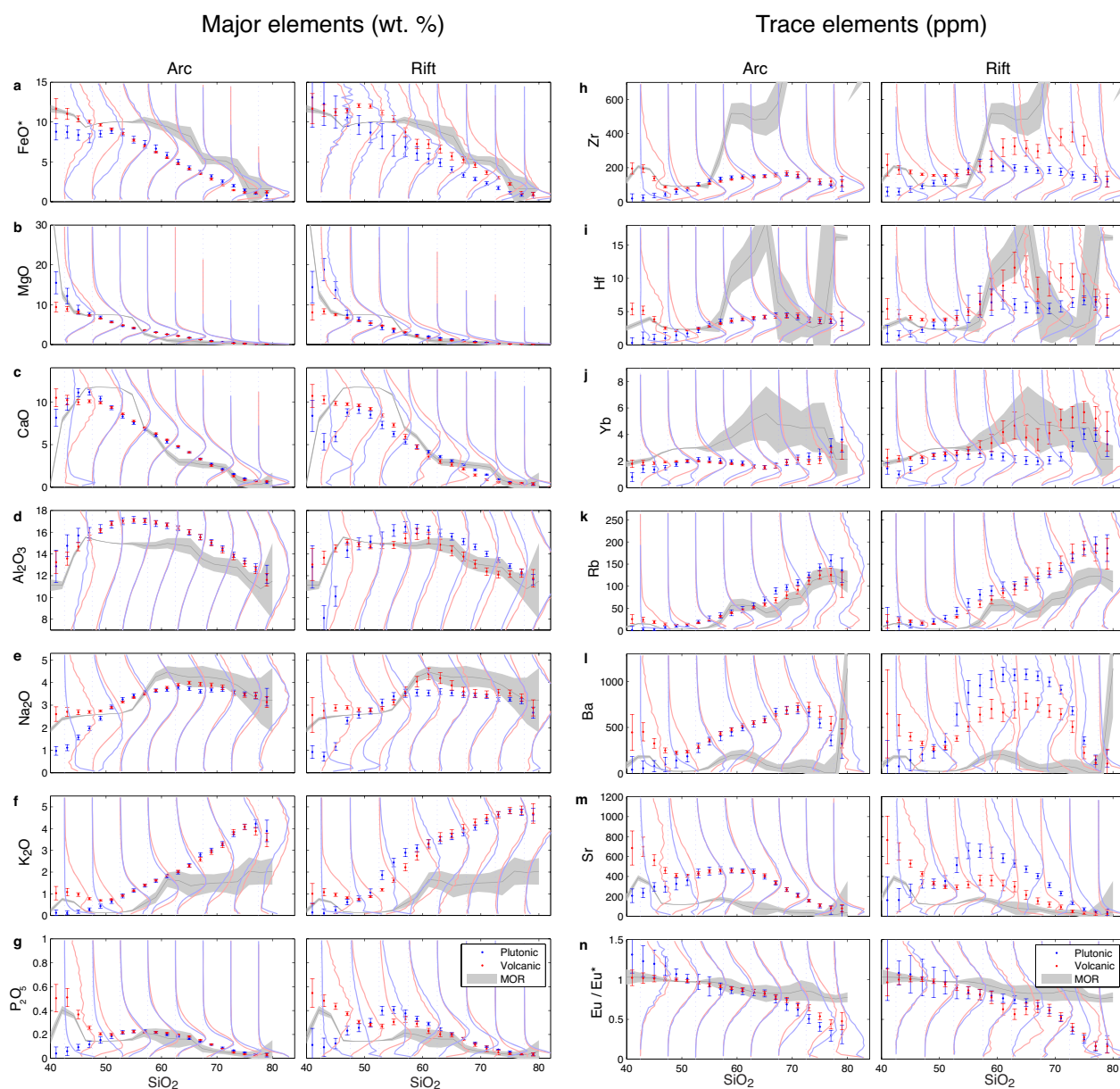
Extended Data Figure 4 | Schematic illustration of instantaneous cumulate compositions required by fractionation trends. a–c, The figure illustrates the compositional offset between melt and cumulate trajectories for the curved trends produced for compatible (a) and incompatible elements (c) during fractional crystallization as well as for a hypothetical linear fractionation trend (b). Instantaneous cumulate compositions will fall along blue lines tangent to the melt curve (red), with the extent of offset along this line correlating with

the efficiency of cumulate extraction. Relatively efficient fractionation, corresponding to a large compositional offset and a high-curvature differentiation trend, is suggested by the observed curvature of major- and trace-element differentiation trends (Figs 1, 2); efficient differentiation is additionally consistent with thermal constraints (for example, ref. 12) which indicate limited heat budgets for crustal differentiation.



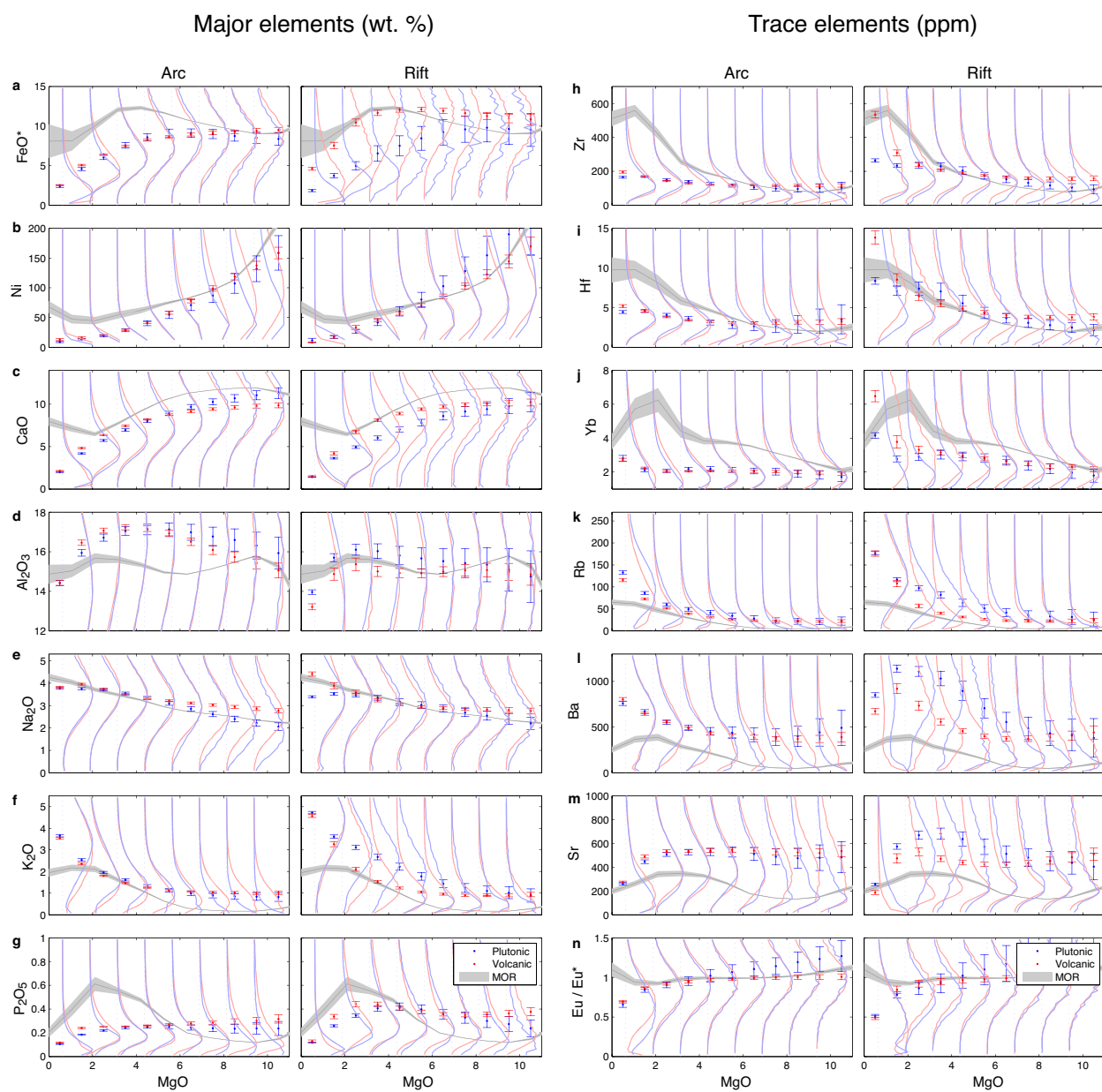
Extended Data Figure 5 | Mean major- and trace-element differentiation trends as a function of SiO_2 . a–g, Major elements; h–n, trace elements. Data are binned in 2 wt% intervals; mean values are shown with 2 s.e.

uncertainties (error bars) for volcanic (red) and plutonic (blue) samples with known ages less than 100 Ma, along with tholeiitic mid ocean ridge differentiates (grey outline).



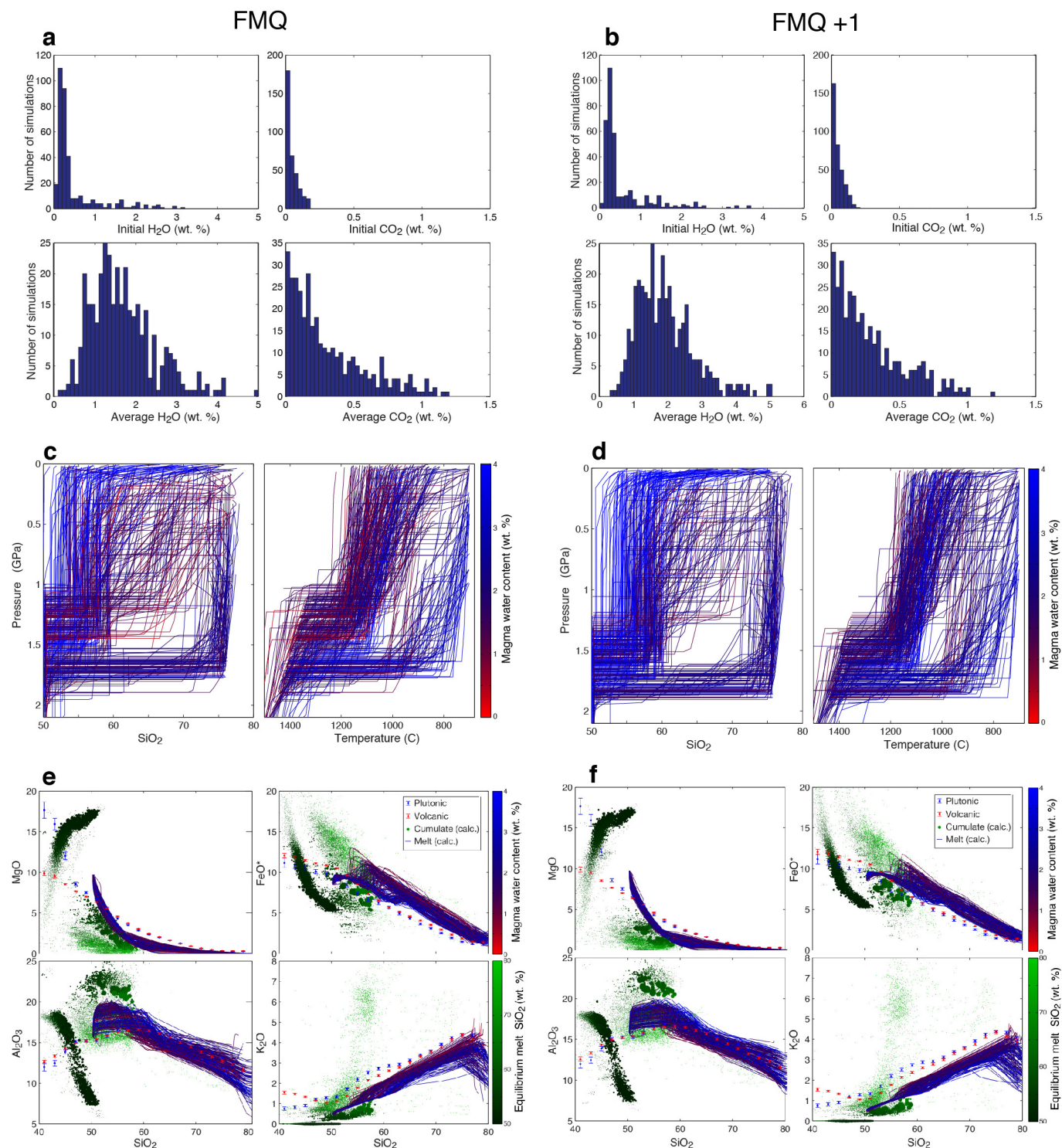
Extended Data Figure 6 | Median major- and trace-element differentiation trends as a function of SiO₂. a–g, Major elements; h–n, (trace) elements. Data (with error bars) are shown with elemental distributions (lines) for volcanic (red) and plutonic (blue) samples, along with tholeiitic mid ocean ridge

differentiates (grey outline). Error bars and MOR outline show median and 2 s.e. uncertainties in 2 wt% SiO₂ bins, while distributions are shown for 5 wt% SiO₂ bins.



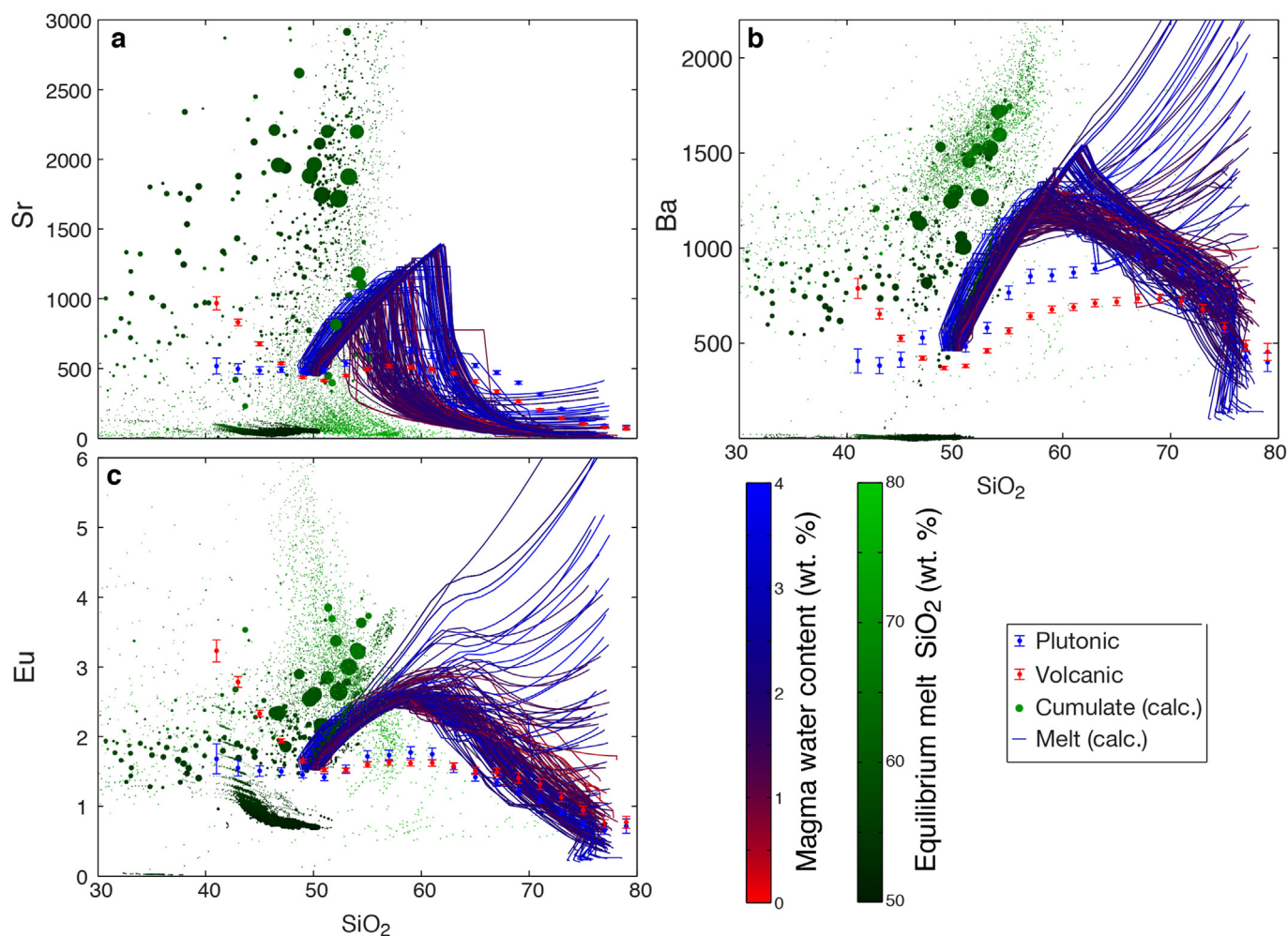
Extended Data Figure 7 | Mean major- and trace-element differentiation as a function of MgO. a–g, Major elements; h–n, trace elements. Data (with error bars) are shown with elemental distributions (lines) for volcanic (red) and

plutonic (blue) samples, along with tholeiitic mid ocean ridge differentiates (grey outline). Error bars and MOR outline show mean and 2 s.e. uncertainties in 1 wt% MgO bins, while distributions are shown for 1.25 wt% MgO bins.



Extended Data Figure 8 | MELTS simulations of magma differentiation by fractional crystallization. The figure shows a comparison of inverted MELTS parameters and compositional paths from $\sim 1.3 \times 10^6$ MELTS simulations with initial oxygen fugacity set by the FMQ buffer (left) versus another $\sim 1.3 \times 10^6$ simulations with initial oxygen fugacity one log unit above the

FMQ buffer (FMQ + 1, right). **a–f**, Initial and average volatile distributions of 350 best-fitting simulations (**a**, **b**); P – T and P – SiO_2 paths of 350 best-fitting MELTS simulations (**c**, **d**); and major-element melt and cumulate compositions for 200 best-fitting MELTS simulations (**e**, **f**).



Extended Data Figure 9 | Trace-element modelling based on MELTS simulation results. The figure shows simulated Sr (a), Ba (b) and Eu (c) differentiation paths (blue lines) and corresponding cumulate compositions (green points) calculated by applying GERM partition coefficients²⁶ to the mineral percentages derived from 200 MELTS simulations with best-fitting

major-element trends. As in Extended Data Fig. 8, best-fitting MELTS simulations were selected out of a total of 1.36×10^6 MELTS simulations with varying initial H_2O , CO_2 , and P - T paths (Methods) and initial oxygen fugacity set by the FMQ buffer. The diameter of each cumulate point is proportional to the volume of cumulates produced at each simulation step.

Extended Data Table 1 | Partition coefficients

	Kd Eu / Kd Eu*	Kd Eu
Orthoclase	69.19	3.22
Anorthite	11.79	1.32
Orthopyroxene	1.043	0.478
Biotite	0.947	0.316
Amphibole	0.8682	1.36
Clinopyroxene	0.8005	3.22
Olivine	0.5560	0.259
Sphene	0.4997	73.7
Allanite	0.4432	124.6
Apatite	0.3927	9.81
Garnet	0.2183	1.28

Mineral/melt partition coefficients (K_d) at 70% SiO₂, interpolated from data of the Geochemical Earth Reference Model K_d Database (adapted from ref. 26, EarthRef.org).

Receptor-mediated exopolysaccharide perception controls bacterial infection

Y. Kawaharada^{1,2}, S. Kelly^{1,2,3}, M. Wibroe Nielsen^{1,2}, C. T. Hjuler^{1,4}, K. Gysel^{1,2}, A. Muszyński⁵, R. W. Carlson⁵, M. B. Thygesen^{1,4}, N. Sandal^{1,2}, M. H. Asmussen^{1,2}, M. Vinther^{1,2}, S. U. Andersen^{1,2}, L. Krusell^{1,2}, S. Thirup^{1,2}, K. J. Jensen^{1,4}, C. W. Ronson^{1,3}, M. Blaise^{1,2†}, S. Radutoiu^{1,2} & J. Stougaard^{1,2}

Surface polysaccharides are important for bacterial interactions with multicellular organisms, and some are virulence factors in pathogens. In the legume–rhizobium symbiosis, bacterial exopolysaccharides (EPS) are essential for the development of infected root nodules. We have identified a gene in *Lotus japonicus*, *Epr3*, encoding a receptor-like kinase that controls this infection. We show that *epr3* mutants are defective in perception of purified EPS, and that EPR3 binds EPS directly and distinguishes compatible and incompatible EPS in bacterial competition studies. Expression of *Epr3* in epidermal cells within the susceptible root zone shows that the protein is involved in bacterial entry, while rhizobial and plant mutant studies suggest that *Epr3* regulates bacterial passage through the plant's epidermal cell layer. Finally, we show that *Epr3* expression is inducible and dependent on host perception of bacterial nodulation (Nod) factors. Plant–bacterial compatibility and bacterial access to legume roots is thus regulated by a two-stage mechanism involving sequential receptor-mediated recognition of Nod factor and EPS signals.

Bacteria display a variety of glycans on their surfaces. Capsular polysaccharides, lipopolysaccharides, cyclic glucans and exopolysaccharides (EPS) form an adaptable dynamic interface involved in, for example, cell-to-cell interactions, immune evasion, pathogenesis, biofilm formation and colonization of ecological niches, such as the intestinal lumen and epithelial surfaces^{1–3}. In the legume–rhizobium symbiosis, mutational analysis of rhizobia has revealed a role for polysaccharides in the chronic intracellular infection of plant cells of nitrogen-fixing root nodules^{4–6}. Legumes inoculated with compatible rhizobia form nodule primordia from new cell divisions in the root cortex in response to Nod factors produced by rhizobia, and simultaneously a bacterial infection process targets these primordia^{6–8}. Infection of most legumes, including the model legume *Lotus japonicus* is initiated when bacteria become entrapped by a curled root hair and divide to form a microcolony. Local cell wall hydrolysis and invagination of the plant plasma membrane then leads to formation of infection threads that are colonized by multiplying bacteria^{9,10}. These tubular structures proceed through the epidermal cell layer into the root cortex before ramifying in the underlying nodule primordia. Finally, rhizobia are released into plant cells in an endocytotic process and maintained in membrane-delimited compartments called symbiosomes.

The role of bacterial EPS in this infection process has been most extensively studied in the *Sinorhizobium meliloti*–alfalfa and *Rhizobium leguminosarum*–pea symbioses. Mutants of both *S. meliloti* and *R. leguminosarum* defective in EPS biosynthesis are severely impaired in their ability to initiate or elongate infection threads^{11,12}. An octasaccharide trimer of *S. meliloti* succinoglycan is symbiotically active¹³. However, the composition and structure of EPS produced by rhizobia varies among strains, suggesting specialized functions in host–rhizobial interactions. Variations in both the monosaccharide composition of subunit backbones and in the degree of polymerization have been reported¹⁴. Exemplifying this variation, some

S. meliloti strains produce both succinoglycan^{15,16} and galactoglucan¹⁷, whereas *R. leguminosarum* EPS is a polymer of a decorated octamer containing glucuronic acid¹⁸. Strain-specific decorations include non-carbohydrate substituents such as O-acetyl, pyruvyl and succinyl groups¹⁴. Illustrating the importance of these variations, *S. meliloti* mutants producing EPS lacking the succinyl groups exhibit impaired infection-thread formation¹².

The precise function of rhizobial EPS and the role of structural variations during infection remain unresolved. An active role in signalling to ameliorate plant defence responses has been proposed^{19,20}, as have passive roles, such as prevention of cellulose-mediated agglutination²¹, sequestering of calcium ions²², and protection against changeable osmotic conditions or against host-derived reactive oxygen species²³. We addressed the question of EPS function by asking whether plants directly perceive EPS in a receptor-mediated process controlling legume–rhizobium compatibility. Here we show that legume host plants monitor the structure of EPS secreted by rhizobia at the initiation of infection and we identify a LysM serine/threonine receptor kinase central to this novel non-self detection mechanism.

The EPR3 exopolysaccharide receptor

Wild-type *Mesorhizobium loti* strain R7A (R7A) synthesizes and secretes an O-acetylated acidic EPS giving rise to mucoid colony growth²⁴. In contrast, R7AexoU mutants of this strain impaired in the *exoU* glucosyltransferase gene secrete a penta-glycan (truncated EPS; Fig. 1a) and have a rough colony appearance²⁴. When inoculated onto wild-type *Lotus japonicus* plants of ecotype Gifu, R7AexoU induced barely visible uninfected root nodule primordia (Fig. 1c, f) compared to infected mature nodules induced by R7A (ref. 24) (Fig. 1b, e). This distinct phenotype provided a well-defined baseline for a suppressor mutant screen aimed at identifying plant genes involved in the response to bacterial EPS. A suppressor line, *exo277*, that developed mature nodules was isolated (Fig. 1d), and

¹Centre for Carbohydrate Recognition and Signalling, Aarhus University, Aarhus 8000 C, Denmark. ²Department of Molecular Biology and Genetics, Aarhus University, Aarhus 8000 C, Denmark.

³Department of Microbiology and Immunology, University of Otago, Dunedin 9054, New Zealand. ⁴Department of Chemistry, University of Copenhagen, Frederiksberg 1871 C, Denmark. ⁵Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30602, USA. [†]Present address: Centre d'études d'agents Pathogènes et Biotechnologies pour la Santé, CNRS, Université de Montpellier, 34293 Montpellier, France.

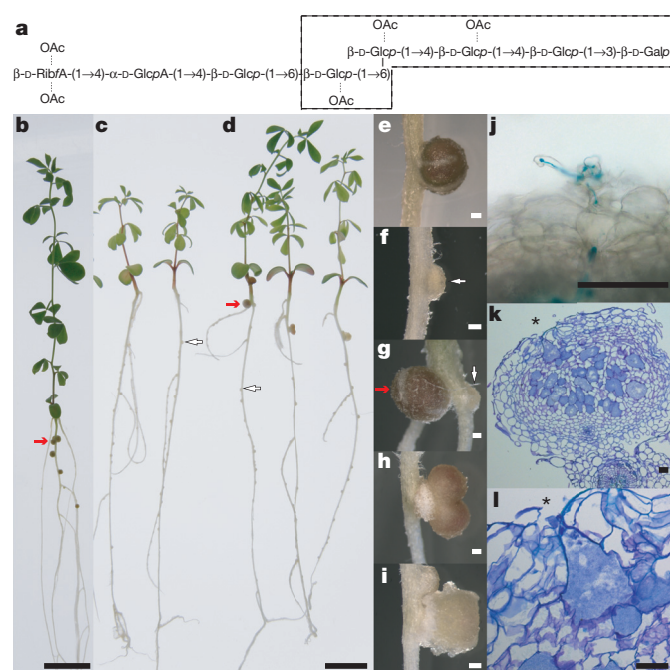


Figure 1 | Phenotype of *epr3* suppressor mutant. **a**, Proposed structure of monomeric EPS (average of three O-acetyl (OAc) groups per monomer). Dashed lines show R7AeoxU truncated EPS. Glcp, glucose; Galp, galactose; RibfA, riburonic acid. **b**, Representative of R7AeoxU-inoculated Gifu. **c**, Two R7AeoxU-inoculated Gifu. **d**, Three R7AeoxU-inoculated exo277/*epr3-10*. **e**, Representative R7AeoxU-inoculated specimens, representative primordium on Gifu. **f**, R7AeoxU-inoculated specimens, representative examples of infected nodules (**g**, **h**) and a large infected white nodule on *epr3-10* (**i**). **b**–**g**, Red and white arrows mark infected and uninfected nodules. **j**, Elongated infection thread of *epr3-10*. **k**, Crack entry in *epr3-10* nodule. **l**, Close up of the infection pocket (indicated by the asterisk). Scale bars: **b**–**d**, 1 cm; **e**–**i**, 20 μm; **j**–**l**, 50 μm. *n* values relating to representative images are provided in Extended Data Table 1.

subsequent inoculation of progeny plants with R7AeoxU constitutively expressing a *DsRed* reporter gene revealed an intermediate frequency of infected nodules (Extended Data Table 1). Both fully developed pink and large white nodules were formed in approximately equal numbers (Fig. 1g–i). In contrast to the small white ‘bumps’, these nodules were infected—as detected by *DsRed* fluorescence. In order to identify the causative gene, the recessive *exo277* locus was genetically mapped to a region on chromosome 2 encompassing a LysM receptor-like kinase gene, *Lys3*²⁵ (Extended Data Fig. 1a). Subsequent sequencing of *Lys3* from *exo277* identified a single-nucleotide mutation (CCC to CTC) that changed Pro to Leu

in the LysM3 domain of the corresponding protein (Extended Data Figs 1b and 2a). In light of the functional characterization presented below, we renamed *Lys3* as *Epr3* (exopolysaccharide receptor 3) and the *exo277* allele as *epr3-10*. To substantiate the gene identification a set of TILLING²⁶ alleles were phenotyped with the R7AeoxU *DsRed* strain (Supplementary Table 1). Two alleles, *epr3-3* and *epr3-9*, had an intermediate frequency of infected nodules similar to *epr3-10* (Extended Data Table 1). A single-nucleotide substitution (TGT to TAT) changed a cysteine in one of the CxC cysteine pairs of the EPR3 ectodomain to tyrosine in *epr3-3* (Extended Data Fig. 2a). In *epr3-9* a splice-site mutation (G to A) at the border of exon 8 led to aberrant splicing (Supplementary Table 1). Additional alleles *epr3-11*, *epr3-12* and *epr3-13*, carrying insertions in exon one, two and three, respectively, were isolated from the LORE1 retrotransposon mutant resource^{27,28} (Extended Data Fig. 1b, Extended Data Table 1 and Supplementary Table 1). Upregulation of *Epr3* expression was not detected in *epr3-11* (Extended Data Fig. 3a) and *epr3-11* was therefore chosen as a presumptive null allele for further phenotypic characterization together with *epr3-10*.

Epr3 consists of ten exons and encodes a protein of approximately 70 kDa. A signal peptide and three LysM motifs are predicted in the amino terminus. The carboxy terminus contains motifs conserved in serine/threonine kinases. Between these domains a single-pass transmembrane segment suggests a topology with LysM motifs in the ectodomain and an intracellular kinase (Fig. 2a and Extended Data Fig. 2b). Supporting this prediction transient expression of an eYFP-tagged EPR3 protein in *Nicotiana benthamiana* leaves localized the protein at the plasma membrane (Extended Data Fig. 4a–c). Phylogenetically EPR3 belongs to the Nod factor receptor 1 (NFR1) protein branch among the 17-member LysM receptor kinase family of *L. japonicus*²⁵. However, both LysM2 and LysM3 have diverged (Extended Data Fig. 5), and EPR3 is unique in the organization of ectodomain motifs. The CxC signature adjacent to the LysM domains of all 17 receptors is located immediately after both LysM1 and LysM2 in EPR3 (Fig. 2b and Extended Data Fig. 2a, b). In contrast to the LysM domains of the CERK1 chitin receptor and NFR1, which have characteristic $\beta 1\alpha 1\alpha 2\beta 2$ structures^{29,30}, $\beta 1\alpha 2\beta 2$ and $\beta 1\alpha 1\beta 2$ configurations are predicted for EPR3 LysM1 and LysM2, respectively. Likewise, motifs involved in the LysM2 chitin-binding site of CERK1 and residues interacting with the N-acetyl of chitin³⁰ are not conserved in EPR3, but are discernable in LysM2 of NFR1 (Fig. 2b). LysM receptor kinases with a similar arrangement of CxC motifs, α -helix predictions and conservation of the Pro substituted in LysM3 of EPR3-10 were found in both dicots and monocots (Extended Data Fig. 2a, b).

EPR3 controls infection at the epidermis

Increased R7AeoxU infection of nodule primordia in *epr3* mutants suggests a role for truncated EPS and EPR3 in regulation of bacterial

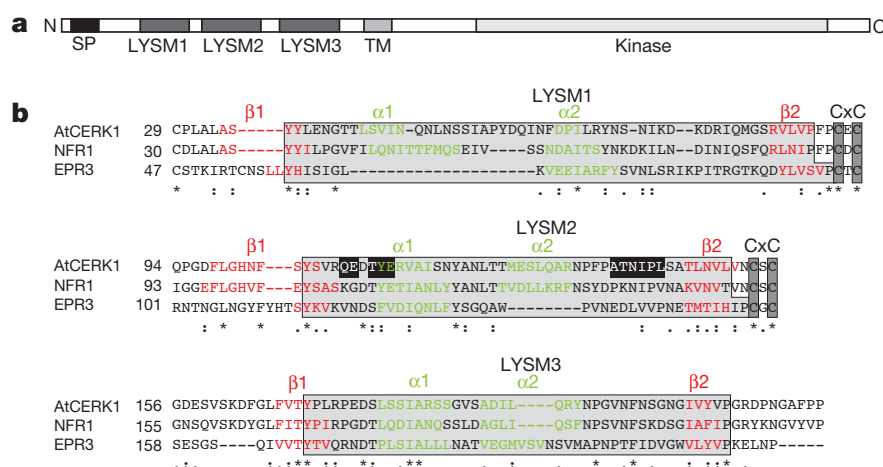


Figure 2 | Domain structure of EPR3 receptor protein. **a**, Diagram presenting the EPR3 domains. SP, signal peptide; TM, transmembrane domain. **b**, Positions of CxC di-cysteine motifs (dark grey), α -helix (green) and β -strand (red) predictions for the LysM domains in CERK1, NFR1 and EPR3. Black boxes indicate CERK1 residues interacting with chitin. Asterisks indicate a conserved amino acid; colons indicate a conservative position; periods indicate a semi-conservative position.

entry across the epidermal cell layer. We therefore recorded the number of infection threads reaching the base of root hair cells at the epidermal–cortical cell boundary (long infection threads; Figs 1j and 3a, b) and/or infection threads not reaching the epidermal–cortical cell boundary (short infection threads; Fig. 3a, c) 10 and 14 days post inoculation (d.p.i.). Inoculation of Gifu with R7AexoU resulted in a five- to tenfold reduction in total infection threads compared to R7A, and at 10 d.p.i. no infection threads reached the base of root hair cells (Fig. 3a). In *epr3-10* and *epr3-11*, infection-thread initiation by R7AexoU was increased and progression was observed (Fig. 3a). At 10 and 14 d.p.i. significantly more long infection threads had reached the cortical boundary in *epr3-10*, and at 14 d.p.i. the total number of infection threads was significantly higher in both *epr3-10* and *epr3-11* (Fig. 3a). Further microscopy of R7AexoU-infected nodules on *epr3-10* and *epr3-11* revealed an alternative infection by crack entry (Fig. 1k, l). These observations suggested a function for *Epr3* in regulation of infection rather than in the nodule organogenic pathway required for cell division and primordia formation. This was further substantiated by the absence of R7AexoU infection in *snf1har1* and *snf2har1* double mutants that develop an excess of spontaneous nodules as a result of autoactivation of the organogenic pathway^{31,32}. Therefore, activation of the host organogenic pathway per se did not rescue the infection deficiency of R7AexoU (Extended Data Table 2a). We conclude that the EPR3 receptor exerts a primary function controlling infection, whether intracellular through infection threads or intercellular via crack entry. In *epr3* mutants this control is defective and R7AexoU is able to initiate more infection threads as well as crack entries, even when the positive effect of EPR3 described below is absent (Fig. 3a).

The suppressor phenotype suggested a ligand–receptor relationship between truncated EPS and EPR3. However, the function of wild-type EPS as a ligand for EPR3 was unclear. The consequences of receptor inactivation and the reciprocal absence of a putative EPS ligand in the R7AexoB mutant were therefore compared. ExoB is the UDP-glucose 4-epimerase that synthesizes UDP-galactose required for the first step of EPS synthesis, and R7AexoB does not synthesize

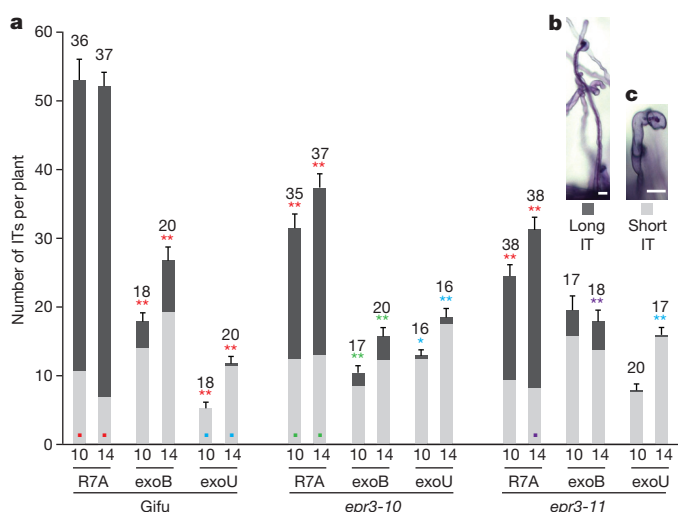


Figure 3 | EPS perception promotes infection thread development.

a, Frequency of infection threads (ITs) after R7A, R7AexoB or R7AexoU inoculation of Gifu, *epr3-10* and *epr3-11*. Summed over ten plants the number of R7AexoU-induced long infection threads at 10 and 14 d.p.i. are: Gifu, 0 and 1.5; *epr3-10*, 5 and 9.4; *epr3-11*, 2 and 3.5. $^{**}P < 0.01$ and $^{*}P < 0.05$ indicate significant differences (*t*-test) in total infection thread number compared to groups at the same d.p.i. and indicated by matching coloured dots. Bars show s.e.m. for the total number of infection threads. See Supplementary Table 2 for extended statistical analysis. *n* values are shown above each column. **b**, **c**, Representative images of long (**b**) and short (**c**) infection threads observed from **a**. Scale bars, 20 μm.

EPS. Unlike R7AexoU, R7AexoB forms infected nodules on Gifu, albeit after a delay²⁴ (Extended Data Fig. 6a).

Infection-thread formation was scored on Gifu, *epr3-10* and *epr3-11* at 10 and 14 d.p.i. with R7AexoB or R7A. The total number of infection threads in Gifu plants inoculated with R7AexoB was reduced to about half at 10 and 14 d.p.i. compared to R7A, suggesting a positive function of wild-type EPS (Fig. 3a). In reciprocal experiments infection-thread formation in R7A-inoculated *epr3-10* and *epr3-11* was reduced to about half, suggesting that EPR3 promotes efficient infection-thread initiation (Fig. 3a). A reduction in infection-thread formation was also observed in R7AexoB compared to R7A-inoculated *epr3-10* and *epr3-11* plants, possibly reflecting a passive physicochemical role of EPS. For both Gifu and the *epr3* mutants the number of infection threads reaching the cortical boundary was significantly reduced at 10 and 14 d.p.i. compared to R7A-inoculated plants.

These observations demonstrate that both wild-type EPS and a functional EPR3 receptor are required for sustained infection-thread initiation. To substantiate a functional relationship between EPS and EPR3 further, the effect of purified EPS (100 μg ml⁻¹) was examined on R7AexoB-inoculated Gifu and *epr3-11* plants. A significant increase in infection threads was observed with Gifu while no increase was detected in *epr3-11* (Fig. 4a). This *Epr3*-dependent positive effect of EPS addition supports the notion of a ligand–receptor relationship.

EPS and truncated EPS are perceived

Based on the inference that truncated EPS causes impaired infection, we sought to define the timing of the host perception and negative reaction. We inoculated Gifu, *epr3-10* and *epr3-11* mutants with R7AexoU and then after 15 min, 1 h, 24 h or 96 h, inoculated these plants with either R7A or R7AexoB. Formation of infected nodules was scored at 14–42 d.p.i. If the second inoculation occurred within 24 h, the prior inoculation with R7AexoU had no effect. This implies that the negative host reaction to truncated EPS is not immediate. However, when the second inoculation was delayed until 96 h after R7AexoU inoculation, an *Epr3*-dependent reduction in nodule numbers was observed (Extended Data Fig. 6b, and control in Extended Data Fig. 6c). Significantly fewer nodules developed on Gifu with R7AexoB compared to R7A as second inoculum. In contrast, this difference in R7AexoU sensitivity between R7A and R7AexoB was absent in *epr3-10* or *epr3-11* mutants. We infer that EPR3 perceives both truncated EPS and wild-type EPS, and that wild-type EPS can out-compete truncated EPS. In support of this interpretation an *Epr3*-dependent increase in infection-thread formation was observed after co-inoculation of R7AexoU with the Nod-factor-deficient but wild-type-EPS-producing R7AΔnodA³³ (Fig. 4b).

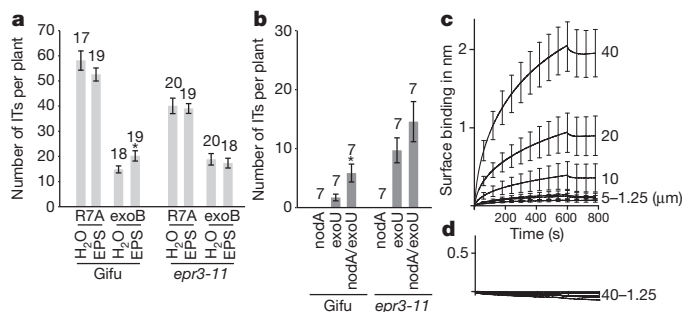


Figure 4 | EPS and truncated EPS are perceived by the EPR3 receptor.

a, Frequency of infection threads (ITs) after addition of purified EPS to R7A or R7AexoB inoculated *epr3-11* and Gifu. No effect of EPS addition was detectable on R7AexoU inoculated plants. **b**, R7AΔnodA and R7AexoU co-inoculation of Gifu and *epr3-11*. $^{*}P < 0.05$ (*t*-test). *n* values are shown above each column. **c**, Sensorgrams show surface binding in nm as a result of binding of increasing concentrations of EPR3 ectodomain to monomeric EPS. **d**, Corresponding sensorgrams for maltohexaose. Values from triplicate measurements; bars show s.e.m.

EPR3 perceives EPS directly

The physical interaction between EPS and EPR3 was assessed in a binding assay using biolayer interferometry technology. EPR3 ectodomain protein was purified following baculovirus expression in insect cells (Extended Data Fig. 7) and octasaccharide EPS monomers were enriched from crude EPS (Extended Data Fig. 8a). EPS monomers were purified by size-exclusion chromatography, verified by mass spectrometry and chemoselectively labelled with biotin by aniline-catalysed oxime formation at the reducing end³⁴ (Extended Data Fig. 8b–e) to enable immobilization with streptavidin biosensors and affinity measurements. Biotin-functionalized maltohexaose was used as carbohydrate control. The sensorgrams obtained with EPS showed time- and EPR3-dependent association and dissociation while no detectable signal was obtained with the maltohexaose control (Fig. 4c, d). A K_d value of $2.7 \pm 0.2 \mu\text{M}$ was determined for binding of monomeric EPS to the EPR3 ectodomain. This *in vitro* assay shows that EPR3 ectodomain can recognize monomeric EPS by direct binding and distinguish maltohexaose.

Cellular localization of *Epr3* expression

Phenotypic characterization suggested EPR3 monitoring of EPS and control of bacterial passage through the root epidermis. *Epr3* promoter activity was therefore investigated in transformed roots of Gifu using both an *pEpr3::GUS* reporter and an *pEpr3::tYFP-NLS* fusion localizing triple YFP protein in the nucleus. GUS activity was detected only in inoculated *pEpr3::GUS* transgenic roots at 3 d.p.i. (Fig. 5a, b). A more detailed study showed *pEpr3::tYFP-NLS* expression in epidermal cells in the susceptible root zone (Fig. 5c–h). A high level of YFP fluorescence was observed in root hairs with progressing infection threads (Fig. 5f–h), supporting a role for EPR3 in the control of epidermal infection.

A two-step process controls infection

In *L. japonicus* the primary signal transduction events activating infection and nodule organogenesis are triggered within minutes following the perception of Nod factors by the NFR1 and NFR5 receptors^{31,35}. To investigate whether EPR3 could substitute the function of NFR1 and/or NFR5 in Nod-factor perception, *Epr3* was expressed in *nfr1-1* and *nfr5-2* mutants. In contrast to *Nfr1* and *Nfr5* expressed from cognate or 35S promoters, full-length *Epr3* (*pEpr3::Epr3*) and *p35S::Epr3* did not restore nodulation (Extended Data Table 2b). To assess if the initial Nod-factor-mediated signalling was perturbed in *epr3* mutants, the expression of two Nod-factor-induced genes, *Nin* and *NfyA1*, was measured following application of 10^{-8} M purified Nod factors (Extended Data Fig. 3c, d). Although *Nin* expression differed between the *epr3-10* mutant that possibly maintains residual kinase activity and the *epr3-11* insertion mutant, expression of both genes was induced in Gifu, *epr3-10* and *epr3-11*, indicating that EPR3 is not essential for Nod-factor perception. Under

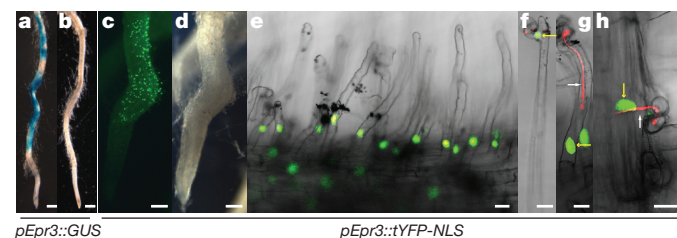


Figure 5 | EPR3 is Nod-factor-induced and expressed in epidermal cells. **a**, Expression of *pEpr3::GUS* in R7A-inoculated root ($n = 25$). **b**, Non-inoculated control ($n = 15$). **c**, **d**, *pEpr3::tYFP-NLS* expression in epidermal cells of the susceptible zone visualized as green fluorescent nuclei, marked with yellow arrows in **f–h** ($n = 20$). **e–h**, *pEpr3::tYFP-NLS* expression in root hairs ($n = 20$). Rhizobia were labelled with DsRed to show infection threads, white arrows. Scale bars: **a–d**, 500 μm ; **e–h**, 20 μm .

the same conditions, *Epr3* expression was Nod-factor-induced in Gifu (Extended Data Fig. 3e). Supporting Nod-factor induction of *Epr3*, no *pEpr3::GUS* activity was detected in roots inoculated with an R7AnodC mutant unable to synthesize Nod factors³³. In contrast, *Epr3* and *Nin* expression was observed after inoculation with R7AexoU or R7AexoB (Extended Data Figs 3a, b and 4d, e).

Discussion

Here we describe a transmembrane receptor kinase, EPR3, and show that rhizobial access to legume roots is controlled by a two-stage mechanism. In *L. japonicus* two LysM receptors, NFR1 and NFR5, with high affinity for Nod factors trigger symbiotic signal transduction^{35,36}. One consequence of this first recognition is *Epr3* expression in root hairs and epidermal cells of the susceptible zone, and we propose that the EPR3 LysM receptor kinase then monitors the EPS status of bacteria during rhizobial infection (Fig. 6). Several lines of evidence support such a role for EPR3: (i) mutation of *Epr3* suppresses the negative effect of truncated EPS; (ii) suppression of truncated EPS effects by normal EPS from R7A Δ nodA is *Epr3*-dependent; (iii) mutation of *Epr3* and absence of EPS ligand in R7AexoB have comparable effects on infection; (iv) inhibition of EPS-deficient R7AexoB by truncated EPS from R7AexoU pre-inoculation is *Epr3*-dependent; (v) *epr3* mutants display defective perception of purified EPS; and (vi) EPR3 ectodomain binds purified monomeric EPS directly. Taken together, our data also suggest that EPR3 distinguishes between EPS variants and acts positively in response to compatible EPS and negatively in response to incompatible EPS (Fig. 6). This explains the reduced infection rate of R7AexoB resulting from absence of a positive response, the positive response to EPS addition, and the abrogated infection of R7AexoU induced by a negative response.

Phenotypic characterization of *epr3* mutants and EPS-addition experiments suggest that plant surveillance of bacterial EPS is exerted at the stage of colonization and infection of epidermal cells, irrespective of whether infection is intracellular through infection threads or intercellular by crack entry. We hypothesize that this recognition

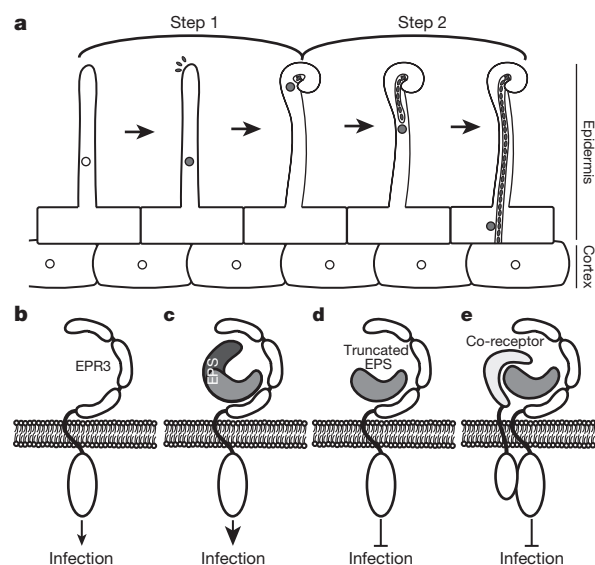


Figure 6 | Model for two-step receptor-mediated recognition of compatible rhizobia. **a**, Step 1: colonization and host-specific initiation of infection triggered by Nod factor binding by NFR1 and NFR5 receptors leads to *Epr3* expression in epidermal cells (grey circles) and root hair curling enclosing rhizobia. Step 2: NFR1 and NFR5 signalling continues and EPR3 perceives EPS. **b–e**, Tentative outcomes from EPR3 perception of compatible EPS and truncated EPS. **b**, A basic level of EPR3 signalling promotes infection. **c**, Perception of compatible EPS increases signalling and promotes persistent infection. **d**, Truncated EPS activates a pathway blocking infection. **e**, Alternatively a co-receptor mechanism inhibits infection.

increases basic signalling to a level required for sustained infection, most likely by amplifying ongoing Nod-factor-mediated signalling and/or suppressing plant defence responses (Fig. 6c and Extended Data Fig. 3b, c). In contrast, perception of incompatible EPS could possibly redirect signalling through a complex with a co-receptor, such as another LysM receptor or a lectin receptor (Fig. 6d, e). In all cases, perception of incompatible EPS may lead to mis-regulated gene activation or activation of a new set of downstream targets^{37,38}. Inactivation of EPR3 in mutants diminishes this change in downstream signalling. As a result, EPR3 surveillance of EPS at the initiation of infection is lifted and infection threads or crack entry are initiated. Wild-type levels of infection are not obtained as *ep3* mutations also eliminate the positive signalling required for sustained infection (Fig. 6c).

Taken together, our data show that EPS serves as a signal perceived by the EPR3 receptor. The ligand specificity, biochemical mechanisms and downstream consequences are a challenge for the future that may add new perspectives on carbohydrate signalling by LysM proteins that are represented across many forms of life, from bacteria to humans^{39,40}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 February 2014; accepted 3 June 2015.

Published online 8 July 2015.

- Chu, H. & Mazmanian, S. K. Innate immune recognition of the microbiota promotes host-microbial symbiosis. *Nature Immunol.* **14**, 668–675 (2013).
- Flemming, H. C. & Wingender, J. The biofilm matrix. *Nature Rev. Microbiol.* **8**, 623–633 (2010).
- Silipo, A. *et al.* Glyco-conjugates as elicitors or suppressors of plant innate immunity. *Glycobiology* **20**, 406–419 (2010).
- Becker, A., Frayse, N. & Sharypova, L. Recent advances in studies on structure and symbiosis-related function of rhizobial K-antigens and lipopolysaccharides. *Mol. Plant Microbe Interact.* **18**, 899–905 (2005).
- Frayse, N., Couderc, F. & Poinot, V. Surface polysaccharide involvement in establishing the rhizobium-legume symbiosis. *Eur. J. Biochem.* **270**, 1365–1380 (2003).
- Jones, K. M., Kobayashi, H., Davies, B. W., Taga, M. E. & Walker, G. C. How rhizobial symbionts invade plants: the *Sinorhizobium-Medicago* model. *Nature Rev. Microbiol.* **5**, 619–633 (2007).
- Desbrosses, G. J. & Stougaard, J. Root nodulation: a paradigm for how plant-microbe symbiosis influences host developmental pathways. *Cell Host Microbe* **10**, 348–358 (2011).
- Oldroyd, G. E. D., Murray, J. D., Poole, P. S. & Downie, J. A. The rules of engagement in the legume-rhizobial symbiosis. *Annu. Rev. Genet.* **45**, 119–144 (2011).
- Fournier, J. *et al.* Mechanism of IT elongation in root hairs of *Medicago truncatula* and dynamic interplay with associated rhizobial colonization. *Plant Physiol.* **148**, 1985–1995 (2008).
- Xie, F. *et al.* Legume pectate lyase required for root infection by rhizobia. *Proc. Natl Acad. Sci. USA* **109**, 633–638 (2012).
- Breedveld, M. W. *et al.* Polysaccharide synthesis in relation to nodulation behavior of *Rhizobium leguminosarum*. *J. Bacteriol.* **175**, 750–757 (1993).
- Cheng, H. P. & Walker, G. C. Succinoglycan is required for initiation and elongation of ITs during nodulation of alfalfa by *Rhizobium meliloti*. *J. Bacteriol.* **180**, 5183–5191 (1998).
- Wang, L. X., Wang, Y., Pellock, B. & Walker, G. C. Structural characterization of the symbiotically important low-molecular-weight succinoglycan of *Sinorhizobium meliloti*. *J. Bacteriol.* **181**, 6788–6796 (1999).
- Skorupska, A., Janczarek, M., Marczak, M., Mazur, A. & Krol, J. Rhizobial exopolysaccharides: genetic control and symbiotic functions. *Microb. Cell Fact.* **5**, 7 (2006).
- Åman, P., Franzen, L. E., Darvill, J. E. & Albersheim, P. Structural elucidation, using HPLC-MS and GLC-MS of the acidic polysaccharide secreted by *Rhizobium meliloti* strain 1021. *Carbohydr. Res.* **95**, 263–282 (1981).
- Reinhold, B. B. *et al.* Detailed structural characterization of succinoglycan, the major exopolysaccharide of *Rhizobium meliloti* Rm1021. *J. Bacteriol.* **176**, 1997–2002 (1994).
- Her, G. R., Glazebrook, J., Walker, G. C. & Reinhold, V. N. Structural studies of a novel exopolysaccharide produced by a mutant of *Rhizobium meliloti* strain Rm1021. *Carbohydr. Res.* **198**, 305–312 (1990).
- Robertson, B. K., Aman, P., Darvill, A. G., McNeil, M. & Albersheim, P. Host-Symbiont Interactions. The structure of acidic extracellular polysaccharides secreted by *Rhizobium leguminosarum* and *Rhizobium trifolii*. *Plant Physiol.* **67**, 389–400 (1981).
- Battisti, L., Lara, J. C. & Leigh, J. A. Specific oligosaccharide form of the *Rhizobium meliloti* exopolysaccharide promotes nodule invasion in alfalfa. *Proc. Natl Acad. Sci. USA* **89**, 5625–5629 (1992).
- Urzaizqui, A. & Walker, G. C. Exogenous suppression of the symbiotic deficiencies of *Rhizobium meliloti* *exo* mutants. *J. Bacteriol.* **174**, 3403–3406 (1992).
- Laus, M. C., van Brussel, A. A. & Kijne, J. W. Role of cellulose fibrils and exopolysaccharides of *Rhizobium leguminosarum* in attachment to and infection of *Vicia sativa* root hairs. *Mol. Plant Microbe Interact.* **18**, 533–538 (2005).
- Aslam, S. N. *et al.* Bacterial polysaccharides suppress induced innate immunity by calcium chelation. *Curr. Biol.* **18**, 1078–1083 (2008).
- Lehman, A. P. & Long, S. R. Exopolysaccharides from *Sinorhizobium meliloti* can protect against H₂O₂-dependent damage. *J. Bacteriol.* **195**, 5362–5369 (2013).
- Kelly, S. J. *et al.* Conditional requirement for exopolysaccharide in the *Mesorhizobium-Lotus* symbiosis. *Mol. Plant Microbe Interact.* **26**, 319–329 (2013).
- Lohmann, G. V. *et al.* Evolution and regulation of the *Lotus japonicus* LysM receptor gene family. *Mol. Plant Microbe Interact.* **23**, 510–521 (2010).
- Perry, J. A. *et al.* A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol.* **131**, 866–871 (2003).
- Fukai, E. *et al.* Establishment of a *Lotus japonicus* gene tagging population using the exon-targeting endogenous retrotransposon LORE1. *Plant J.* **69**, 720–730 (2012).
- Urbański, D. F., Malolepszy, A., Stougaard, J. & Andersen, S. U. Genome-wide LORE1 retrotransposon mutagenesis and high-throughput insertion detection in *Lotus japonicus*. *Plant J.* **69**, 731–741 (2012).
- Tanaka, K. *et al.* Role of LysM receptors in chitin-triggered plant innate immunity. *Plant Signal. Behav.* **8**, e22598 (2013).
- Liu, T. *et al.* Chitin-induced dimerization activates a plant immune receptor. *Science* **336**, 1160–1164 (2012).
- Madsen, L. H. *et al.* The molecular network governing nodule organogenesis and infection in the model legume *Lotus japonicus*. *Nature Commun.* **1**, 10 (2010).
- Tirichine, L. *et al.* spontaneous root-nodule formation in the model legume *Lotus japonicus*: a novel class of mutants nodulates in the absence of rhizobia. *Mol. Plant Microbe Interact.* **19**, 373–382 (2006).
- Rodpohong, P. *et al.* Nodulation gene mutants of *Mesorhizobium loti* R7A-nodZ and *nolL* mutants have host-specific phenotypes on *Lotus* spp. *Mol. Plant Microbe Interact.* **22**, 1546–1554 (2009).
- Thygesen, M. B. *et al.* Nucleophilic catalysis of carbohydrate oxime formation by anilines. *J. Org. Chem.* **75**, 1752–1755 (2010).
- Radutoiu, S. *et al.* Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**, 585–592 (2003).
- Brogammer, A. *et al.* Legume receptors perceive the rhizobial lipochitin oligosaccharide signal molecules by direct binding. *Proc. Natl Acad. Sci. USA* **109**, 13859–13864 (2012).
- Höglund, N. *et al.* Dissection of symbiosis and organ development by integrated transcriptome analysis of *Lotus japonicus* mutant and wild-type plants. *PLoS ONE* **4**, e6556 (2009).
- Jones, K. M. *et al.* Differential response of the plant *Medicago truncatula* to its symbiont *Sinorhizobium meliloti* or an exopolysaccharide-deficient mutant. *Proc. Natl Acad. Sci. USA* **105**, 704–709 (2008).
- Laroche, F. J. *et al.* The embryonic expression patterns of zebrafish genes encoding LysM-domains. *Gene Expr. Patterns* **13**, 212–224 (2013).
- Wong, J. E. M. *et al.* An intermolecular binding mechanism involving multiple LysM domains mediates carbohydrate recognition by an endopeptidase. *Acta Crystallogr. D* **71**, 592–605 (2015).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Danish National Research Foundation grant no. DNRF79 and the ERC Advanced Grant 268523. We thank S. Bucholdt for technical help with *Epr3* promoter-GUS fusions, T. Brock-Nannestad for technical assistance with FT-ICR mass spectrometry and J. Sullivan for comments on the manuscript. A.M. and R.W.C. were supported by the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, US Department of Energy grant (DE-FG02-93ER20097) to the DOE Center for Plant and Microbial Complex Carbohydrates at the Complex Carbohydrate Research Center.

Author Contributions Y.K., identification of EMS and LORE1 *ep3* mutations, characterization of the *Epr3* gene, and phenotypic characterization; S.K., phenotypic characterization; M.W.N., identification of TILLING *ep3* mutations, characterization of the *Epr3* gene, complementation experiments; C.T.H. and M.B.T., biotin tagging of EPS and maltohexaose and affinity studies; K.G. and M.V., purification of EPR3 ectodomain and affinity studies; A.M., R.W.C. and C.T.H., purification and characterisation of monomeric EPS; N.S., screening for suppressor mutants and isolation of the *ep3-10* allele; M.H.A., technical assistance; S.U.A., analysis of next-generation sequencing data; L.K., localization in *N. benthamiana*; K.J.J., S.T. and C.W.R. conceived experiments; M.B. and S.R. conceived and coordinated experiments; J.S. conceived experiments and coordinated. JS wrote the manuscript with input from Y.K., S.K., M.W.N., M.B., C.T.H., M.B.T., A.M., C.W.R. and S.R.

Author Information The *Epr3* gene sequence has been deposited in GenBank under accession number AB506700.1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S. (stougaard@mbg.au.dk).

METHODS

Plant material and growth conditions. Seeds were treated with concentrated sulfuric acid, surface sterilized in 0.5% bleach and germinated on wet filter paper for 3 days as previously described⁴¹. Seeds with emerging radicles were transferred to square Petri dishes with agar slopes (plates) or magenta boxes (Sigma-Aldrich) containing 0.25 × B&D medium⁴². The plate medium was solidified with 1.4% Agar Noble (Difco) and the surface of the slope was covered with filter paper (AGF 651, Frisenette ApS). A metal bar with 3-mm holes for roots was inserted at the top of the agar slope. Plates were incubated in an upright position in specially fabricated boxes, excluding light from the roots below the metal bar. Magenta boxes containing lightweight expanded clay aggregate (LECA, Saint-Gobain Weber A/S, <http://www.weber.dk>, 2–4 mm diameter) and vermiculite size M (Damolin A/S, <http://www.damolin.dk>) in a 4:1 mixture were supplemented with 60 ml 0.25 × B&D. Plants were incubated at 21 °C under a 16 h light/8 h dark cycle. For nodulation studies plates containing ten plants each were inoculated with 500–1,000 µl of *M. loti* at an optical density of OD₆₀₀ = 0.02–0.05.

Bacterial strains and growth conditions. Wild-type *M. loti* strain R7A^{43,44} and strains R7AexoU, R7AexoB, R7AΔnodA and R7AΔnodC derived from R7A were grown in TY/YMB medium at 28 °C. The corresponding strains constitutively expressing *DsRed* or *lacZ* were obtained through the introduction of plasmids pSKDSRED²⁴ or pXLGD4⁴⁵, respectively, and were used for the visualization of infection threads. Roots used for counting infection threads were inspected using a Zeiss Axioplan 2 image fluorescence microscope. Bacteria expressing *lacZ* within infection threads were visualized as previously described⁴⁶. Infection threads were also visualized following staining of roots in 0.01% toluidine blue for 1 min and rinsing in several changes of H₂O. *Agrobacterium rhizogenes* strains AR12 or AR1193^{47,48} were used for all hairy root transformation experiments according to previously described methods⁴⁹. The following concentrations of antibiotics were used for *M. loti* and *A. rhizogenes* strains: ampicillin, 100 µg ml⁻¹; tetracycline, 2–10 µg ml⁻¹; gentamicin, 50 µg ml⁻¹; rifampicin, 100 µg ml⁻¹; spectinomycin, 100 µg ml⁻¹; neomycin, 100 µg ml⁻¹.

Suppressor screen. Seeds from *L. japonicus* ecotype Gifu⁴¹ were treated with ethyl methanesulfonate (EMS) and N-ethyl-N-nitrosourea (ENU) and M2 plants of self-fertilized M1 plants were inoculated with the R7AexoU strain. In the primary screen, 500 to 1,000 plants originating from ten M1 seeds pools were grown in autoclaved trays (35 × 55 × 25 cm) containing LECA and vermiculite in a 3:1 mix and 5 l 0.25 × B&D for 6 to 8 weeks in greenhouse conditions. To avoid flooding the plant roots a metal plate covered with cloth was placed in the bottom of the trays. Based on nodulation and growth phenotypes approximately 500 M2 plants from 1,246 M1 pools were identified as putative suppressor mutants and transferred to soil for seed production. Subsequently, progeny from these putative suppressor mutants were tested for inheritance of the observed phenotype using the square plate setup described above at 21 °C in growth cabinets. This procedure identified the exo277 (*ep3-10*) suppressor mutant line. Finally an *ep3* allelic mutant series was established from lines obtained from the Lotus TILLING facility (RevGenUK, <http://revgenuk.jic.ac.uk/>) and from the LORE1 insertional mutagenesis project (<http://users-mb.au.dk/pmgrp/index.html>). All mutants were genotyped using the primers listed in Supplementary Table 1.

Genetic analysis. The mutation carried by the exo277 (*ep3-10*) suppressor line was mapped using a F2 population established from a cross to *L. japonicus* ecotype MG113 obtained from Legume Base (<http://www.legumebase.brc.miyazaki-u.ac.jp/>). Marker information was obtained from deep sequencing of the MG113 genome and comparison towards the Gifu genome sequence. A list of SSR, indel and SNP markers is shown in Supplementary Table 3.

Protein analysis. BlastP (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to find EPR3 homologues from other plants. Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) was used for multiple sequence alignment. The phylogenetic analysis was performed using the CLC Main Workbench v6.5. SignalP 4.1 Server (<http://www.cbs.dtu.dk/services/SignalP/>) and TMHMM Server v2.0 (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>) predicted signal peptides and transmembrane domains. PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) was used for secondary structure prediction.

Expression and purification of the EPR3 ectodomain. The *Epr3* ectodomain nucleotide sequence was codon-optimized for insect cell expression (GeneScript, Piscataway) and engineered with an N-terminal gp67 signal peptide and a C-terminal hexahistidine tag. The insert was cloned into the transfer vector pOET2 (Oxford Expression Technologies) and recombinant baculoviruses were obtained with the flashBAC GOLD system (OET) according to the manufacturer's instructions. Sf9 cells (1.5 × 10⁶ cells per ml) cultured at 26 °C in suspension in serum-free HyClone SFX-Insect medium (FisherScientific) supplemented with 100 U l⁻¹ penicillin/streptomycin (Life technologies) and 1% (v/v) chemically defined lipid concentrate (Gibco) were infected with recombinant passage 3 virus particles (20 ml per l of culture). After expression for 5 days, medium

supernatant was harvested by centrifugation and was subjected to an overnight dialysis step against 50 mM Tris-HCl pH 8.0, 200 mM NaCl at room temperature.

EPR3 ectodomain was purified from the medium by Ni-IMAC using a HisTrap excel affinity column (GE Healthcare). The protein was eluted with 50 mM Tris-HCl, 200 mM NaCl and 500 mM imidazole. After dialysis against 50 mM Tris-HCl pH 8.0, 200 mM NaCl, EPR3 ectodomain was further purified by another step of Ni-IMAC using a HisTrap HP affinity column (GE Healthcare). After elution and subsequent concentration, the purification was finalized by a SEC step on a Superdex 75 10/300 GL column connected to an ÄKTA Purifier system (GE Healthcare) in a buffer containing PBS, pH 7.2. The protein is expressed in different glycosylation states, visible as distinct bands on an SDS-PAGE gel.

Purification of monomeric exopolysaccharide. Monomeric EPS was recovered from minimal G/RDM culture media by sequential precipitation with absolute EtOH. Initially high molecular weight polysaccharides were precipitated with three volumes of EtOH (v/v) as previously described²⁴. The remaining supernatant was concentrated down by rotary evaporation and the low-molecular-weight oligosaccharides were recovered from the solution by additional precipitation with nine volumes of EtOH (v/v).

Crude nine-volume EtOH precipitate was dialysed against deionized H₂O using SpectraPor 1000MWCO tubing, at 4 °C. Concentrated dialysate was filtered through nylon centrifuge filters (0.22 µm pore size) and fractionated on Superdex Peptide 10/300 GL column (GE Healthcare Life Sciences), assembled with Agilent Technologies 1200 Series system, with 50 mM ammonium acetate buffer pH 6.7 (BioXtra, ≥98%, Sigma Aldrich) used as the eluent. The 48 min separation was carried out in isocratic conditions, and with 0.5 ml min⁻¹ column flow, at room temperature. Eluting fractions were simultaneously evaluated with a refractive index detector (RID-10A, Shimadzu) and multiple wavelength detector (Agilent Technologies 1200 Series) at λ = 210, 230 and 280 nm, respectively. The fractions giving a positive response in RID were pooled, frozen and ammonium acetate buffer was evacuated by multiple freeze-drying cycles. The quality of monomeric EPS was evaluated by MALDI mass spectrometry and by NMR analysis (A. Muszyński, personal communication, 2015).

MALDI-TOF mass spectrometry analysis of monomeric polysaccharide. Monomeric EPS fraction was dissolved in water and mixed 1:1 (v/v) with either 0.5 M Super-DHB (9:1 mixture of DHB and 2-hydroxy-5-methoxybenzoic acid) or 0.5 M THAP (2,4,6-trihydroxyacetophenone monohydrate) matrix in methanol and spotted onto a stainless steel target plate. MALDI spectra were recorded in the negative reflector ionization mode using Applied Biosystem's AB SCIEX TOF/TOF 5800 system.

Synthesis of EPS octaose-biotin and maltohexaose-biotin conjugates. Carbohydrates were conjugated to biotin following an approach analogous to the one described by Broghammer *et al.*³⁶. Heterobifunctional aminoxy-thiol OEG linker, O-(2-(2-(2-(tritylsulfanyloxy)ethoxy)ethoxy)ethyl)hydroxylamine, was prepared as described previously⁵⁰. All other chemicals were purchased from Sigma-Aldrich, and used without further purification. Liquid chromatography-mass spectrometry (LCMS) was performed on an UltiMate 3000 UHPLC⁺ focused system, fitted with a Thermo Fisher MSQ Plus ESI MS, using a Phenomenex Gemini 5 µm, C18, 110 Å, 50 × 4.6 mm analytical column, and using a gradient of 5–100% acetonitrile in water containing 0.1% formic acid. Preparative high-performance liquid chromatography (HPLC) was performed on an UltiMate 3000 instrument fitted with a Waters 996 photodiode detector, using a Phenomenex Luna 5 µm, C18(2), 100 Å, 250 × 100 mm semi-preparative column. Both conjugates were purified using a gradient of 5–60% acetonitrile in water containing 0.1% formic acid, and with a flow rate of 5 ml min⁻¹. High-resolution mass spectrometry (HRMS) was performed on a Bruker Solarix XR ESI/MALDI-FT-ICR-MS instrument equipped with a 7 T magnet. The instrument was run in ESI-mode and externally calibrated with NaOTFA cluster ions. Data was processed using Bruker Data Analysis 4.1. Purified EPS octaose (1.7 mg, 1.18 µmol, 5 mM) was dissolved in 50 mM NaOAc buffer containing 50 mM aniline in MeCN-H₂O (1:1), pH 4.5, and aminoxy-thiol OEG linker (10 mM) was added³⁴. The resulting mixture was allowed to react at 40 °C for 16 h. Conjugate oxime formation was confirmed by LCMS. MS (ESI) calculated for [M+H, 3Ac]⁺ = 1,888.6, found 1,889.0, and calculated for [M+H, 2Ac]⁺ = 1,846.6, found 1,846.9. The mixture was concentrated under a N₂ flow, and triturated three times with *n*-heptane to remove excess linker. To the residue was added a 5% solution of triethylsilane in trifluoroacetic acid (5 ml). After 15 min at room temperature the mixture was concentrated under a N₂ flow and triturated three times with *n*-heptane. The residue was dissolved in 50 mM Na₂B₄O₇ buffer, pH 8.3, to a concentration of EPS of 2.5 mM, upon which N-(2-(2-(2-(biotinylamido)ethoxy)ethoxy)ethyl)-2-iodoacetamide (5 mM) was added³⁰. The reaction was conducted at room temperature for 15 min, after which the product was purified by preparative HPLC. The chromatogram displayed several overlapping peaks, owing to microheterogeneity in the distribution of

OAc groups, ranging from one to three OAc groups. The EPS octamer–biotin conjugate (27% yield) was quantified using the HABA–avidin biotin–quantification kit from Pierce. HRMS (FT-ICR) calculated for $[M+H, 3Ac]^+ = 2,060.7067$, found 2,060.7066, and calculated for $[M+H, 2Ac]^+ = 2,018.6961$, found 2,018.7004, and calculated for $[M+H, 1Ac]^+ = 1,976.6856$, found 1,976.6888, and calculated for $[M+2H, 3Ac]^{2+} = 1,030.8574$, found 1,030.8559, and calculated for $[M+2H, 2Ac]^{2+} = 1,009.8521$, found 1,009.8514, and calculated for $[M+2H, 1Ac]^{2+} = 988.8468$, found 988.8463.

The synthesis of a maltohexose–biotin conjugate was performed essentially as for the EPS–biotin conjugate. HRMS (FT-ICR) calculated for $[M+H]^+ = 1,612.6252$, found 1,612.6214.

Biolayer interferometry binding studies. EPR3 ectodomain interaction with EPS octamer was measured on an Octet RED biolayer interferometer (Pall ForteBio) that measures wavelength shift ($\Delta\lambda$) of reflected white light in real time. Binding of a protein to immobilized ligand⁵¹ will change the optical thickness of a biosensor resulting in a quantifiable wavelength shift (see <http://www.fortebio.com/bli-technology.html>). Biotinylated EPS octase, and maltohexaose as control, were immobilized on streptavidin biosensors (Pall ForteBio) at a concentration of 250 nM for 5 min. Immobilization levels were identical for the EPS and maltohexaose conjugates, amounting to approximately 0.6 nm and 0.5 nm, respectively. Interaction with EPR3 ectodomain was measured in dilution series at protein concentrations ranging from 1.25–40 μ M in phosphate buffered saline, pH 7.0, with 0.01% Tween 20, at 30 °C for 10 min. Subsequently, dissociation was conducted in protein-free phosphate buffered saline, pH 7.0, with 0.01% Tween 20, at 30 °C for 3 min. Background measurements using biosensors immobilized with free biotin were subtracted from EPS and maltohexaose data in order to account for unspecific binding. Sensorgrams were aligned to the baseline using ForteBio Data Analysis 7.0 (Pall ForteBio), and data processing with GraphPad PRISM 6.05 software yielded a $K_d = 2.7 \pm 0.2 \mu$ M from triplicate measurements with EPR3 ectodomain.

Expression analysis. For expression analysis in roots, plants were grown and Nod factors applied modifying a method previously described⁵². mRNA was isolated from Nod-factor-treated (10^{-8} M) roots using Dynabeads mRNA DIRECT kit (Invitrogen). RevertAid M-MuLV Reverse Transcriptase (Fermentas) was used for cDNA synthesis. All cDNA samples were tested for genomic DNA contamination using primers specific for the *Nin* gene promoter²⁵. A Lightcycler480 instrument and Lightcycler480 SYBR Green I master (Roche Diagnostics GmbH) was used for the real time quantitative PCR. ATP-synthase (*ATP*), ubiquitin-conjugating enzyme (*UBC*) and protein phosphatase 2A (*PP2A*) were used as reference genes⁵³. Melting curve analyses and sequencing of all the amplified products were performed to verify the specificity of the primers. The relative quantification software (Roche) was used to calculate the normalized efficiency-corrected relative transcript levels. The geometric mean of the relative transcript levels for the three biological (each consisting of ten plants) and three technical repetitions and the corresponding upper and lower 95% confidence were calculated⁵⁴. *Epr3*, *Nin*, *NfyA1* and reference genes primers are listed in Supplementary Table 4.

Transformation of *Nicotiana benthamiana* leaves. The *Epr3::YFP-HA* construct was obtained by amplifying the *Epr3* cDNA sequence (without the stop codon) with primers containing *attB* sites and a thrombin site following the *attB2* site in the reverse primer. The amplified product was recombined into the *pDONR207* vector (Invitrogen) in a Gateway BP reaction (Invitrogen), which further recombined with the vector *pEarleygate101*⁵⁵ in the Gateway LR reaction (Invitrogen) to construct *pEarleygate::p35S::Epr3::YFP-HA*.

A. tumefaciens strain AGL1 transformed with the *pEarleygate::p35S::Epr3::YFP-HA* expression clone was grown to high density in liquid culture, centrifuged at 1300g and suspended in 10 mM Mes-KOH, 10 mM MgCl₂ and 0.15 mM acetosyringone. The culture was diluted to OD₆₀₀ = 0.2 and mixed with a 1/10 volume of an *A. tumefaciens* culture expressing the silencing suppressor p19⁵⁶. Leaves of 3–4-week-old *N. benthamiana* plants were infiltrated with the mix and analysed at 3 days post infiltration using a Zeiss LSM510 META confocal microscope. For co-localization of EPR3–eYFP with the plasma membrane the membrane dye FM 4-64 (Invitrogen) was used.

Promoter GUS and tYFP-NLS analysis. A putative *Epr3* promoter region of 1,016 bp was amplified from *L. japonicus* Gifu genomic DNA by PCR using primers containing *attB* sites and recombined into *pDONR207* (Invitrogen) using the Gateway BP reaction (Invitrogen) to create the entry clone *pDONR207::pEpr3*. The entry clone was recombined using Gateway LR reactions (Invitrogen) with destination vectors *pIV10::GW::GUS* and *pIV10::GW::tYFP-NLS* (D. Reid, personal communication, 2014) to create the *pIV10::pEpr3::GUS* and *pIV10::pEpr3::tYFP-NLS* constructs. The plasmids containing these promoter fusions were transformed into *A. rhizogenes*.

The *pIV10::pEpr3::GUS* and *pIV10::pEpr3::tYFP-NLS* constructs in *A. rhizogenes* were transformed into *L. japonicus* as previously described⁴⁷. Transformed plants were transferred to Magenta boxes and grown for a week before inoculation with 1,000 μ l of a dilution (OD₆₀₀ = 0.01–0.02) of *M. loti* strain R7A DsRed or MAFF303099 DsRed per plant. For *pEpr3::tYFP-NLS* analysis, transgenic roots were observed for nuclear fluorescence using a Zeiss LSM510 META confocal microscope. For *pEpr3::GUS* and *pNin::GUS* analysis, roots were stained using GUS-staining buffer (0.05 M sodium phosphate buffer pH 7.0, 10 mM EDTA, 20% methanol (v/v), 50 mg ml⁻¹ X-gluc dissolved in DMSO) at 37 °C overnight. Stained roots were washed with 70% (v/v) ethanol and stored at 4 °C. Root and nodule sections used for visualizing the spatial GUS activity were obtained using a Leica M165 FC stereomicroscope.

Nodule sectioning. Two- to six-week-old nodules induced by *M. loti lacZ* were fixed with 1.25% glutaraldehyde in 0.1 M phosphate buffer, and then stained in LacZ reaction buffer (0.1 M sodium phosphate buffer pH 7.0, 0.64 mg ml⁻¹ X-gal dissolved in *N*-methyl-2-pyrrolidone, 5 mM potassium ferrocyanide, 5 mM potassium ferricyanide, 0.1% (w/v)) at room temp overnight. The nodule samples were embedded in 3% agarose in water, sliced in thick layer sections (50 to 100 μ m) using a Leica VT 1000S vibratome. For semi-thin layer sections, nodules were fixed in a mixture of 4% (w/v) paraformaldehyde and 1% glutaraldehyde in 0.1 M phosphate buffer pH 7.2. Samples were then embedded using a Technovit 7100 (Heraeus Kulzer), sliced in semi-thin layer sections (7 μ m) using a Leica RM2045 microtome, and then mounted onto microscope slides. Sections were stained with 0.1% toluidine blue in water and observed using a Zeiss Axioplan 2 image fluorescence microscope.

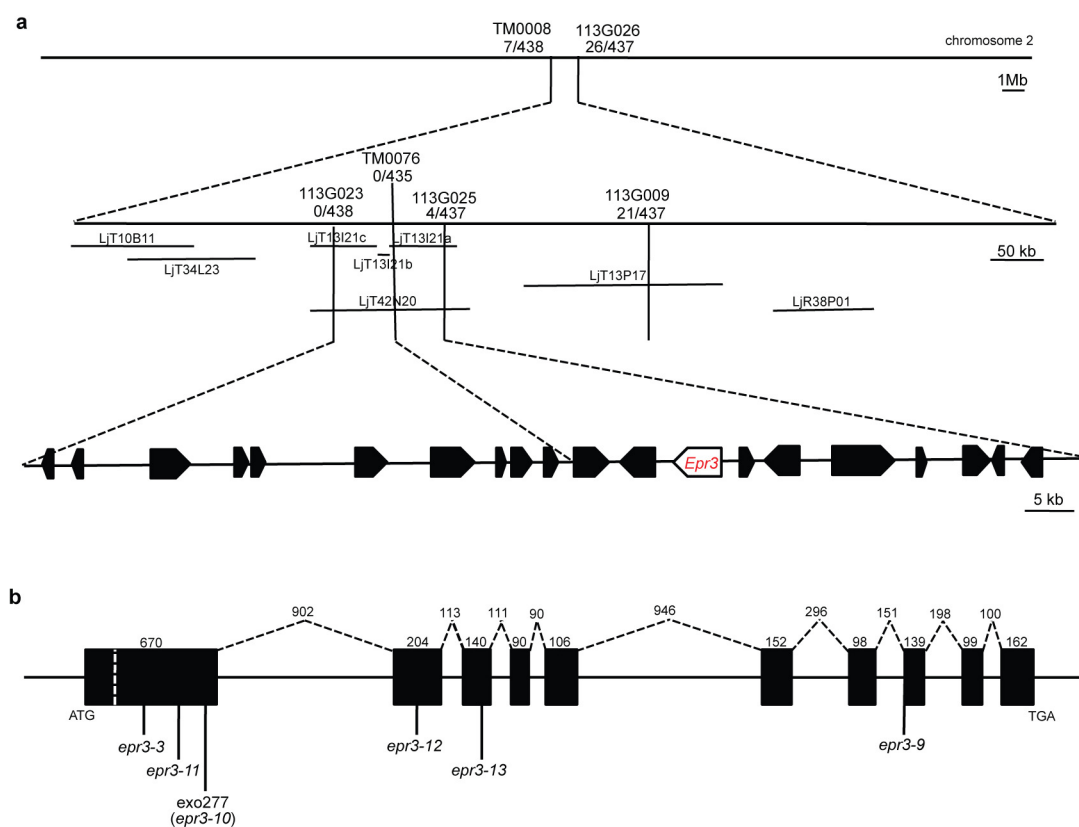
Competition analysis. *L. japonicus* Gifu seedlings, on square growth plates, were inoculated with 1 ml of an OD₆₀₀ = 0.02 R7AexoU suspension per plate and grown at 21 °C for 15 min, 1 h, 24 h and 96 h before subsequent inoculation with 1 ml of an OD₆₀₀ = 0.02 R7A or R7AexoB suspension. In control experiments the symbiotically impaired Nod factor mutant R7A Δ nodA was used as the initial inoculum. A 1:1 mixture of R7AexoU and R7A Δ nodA was used in the coinoculation study (Fig. 4b and Extended Data Fig. 6). For the EPS treatment experiment, Gifu and *epr3-11* seedlings were inoculated with 1 ml R7AexoB DsRed (OD₆₀₀ = 0.02) per plate. After 72 h, 100 μ g ml⁻¹ of LMW EPS purified from R7A Δ ndvB or sterile water was added. Plants were grown at 21 °C for 7 days, and infection threads were counted.

Constructs for complementation of *nfr1* and *nfr5* mutants. A full-length genomic *Epr3* construct containing 1,016 bp upstream from the start codon and 2,014 bp downstream of the stop codon was amplified by PCR and cloned into *pENTER* vector (Invitrogen), to construct the entry clone named *pENTER::pEpr3::Epr3*. The entry clone was recombined using Gateway LR reactions (Invitrogen) with destination vectors named *pIV10::GW*, to construct *pIV10::pEpr3::Epr3*. The construct *p35S::Epr3::t35S* was obtained by amplifying the *Epr3* cDNA sequence from start to stop codon with primers containing *attB* sites and recombining into the *pDONR207* vector (Invitrogen) which was further recombined with the destination vector *pIV10::p35S::GW::t35S* in a Gateway LR reaction (Invitrogen) creating the expression clone *pIV10::p35S::Epr3::t35S*.

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

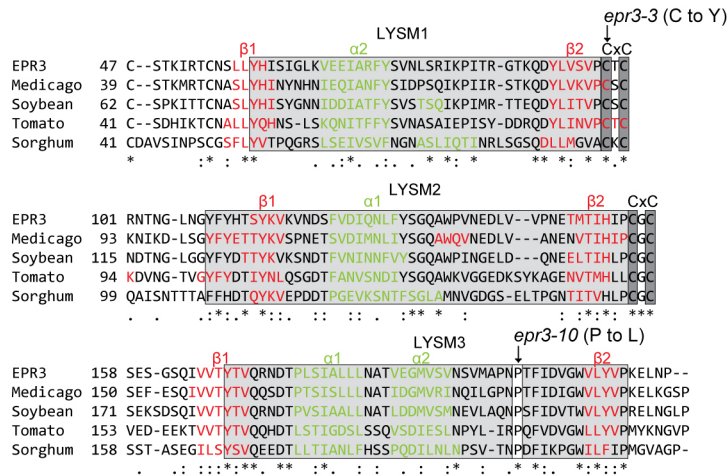
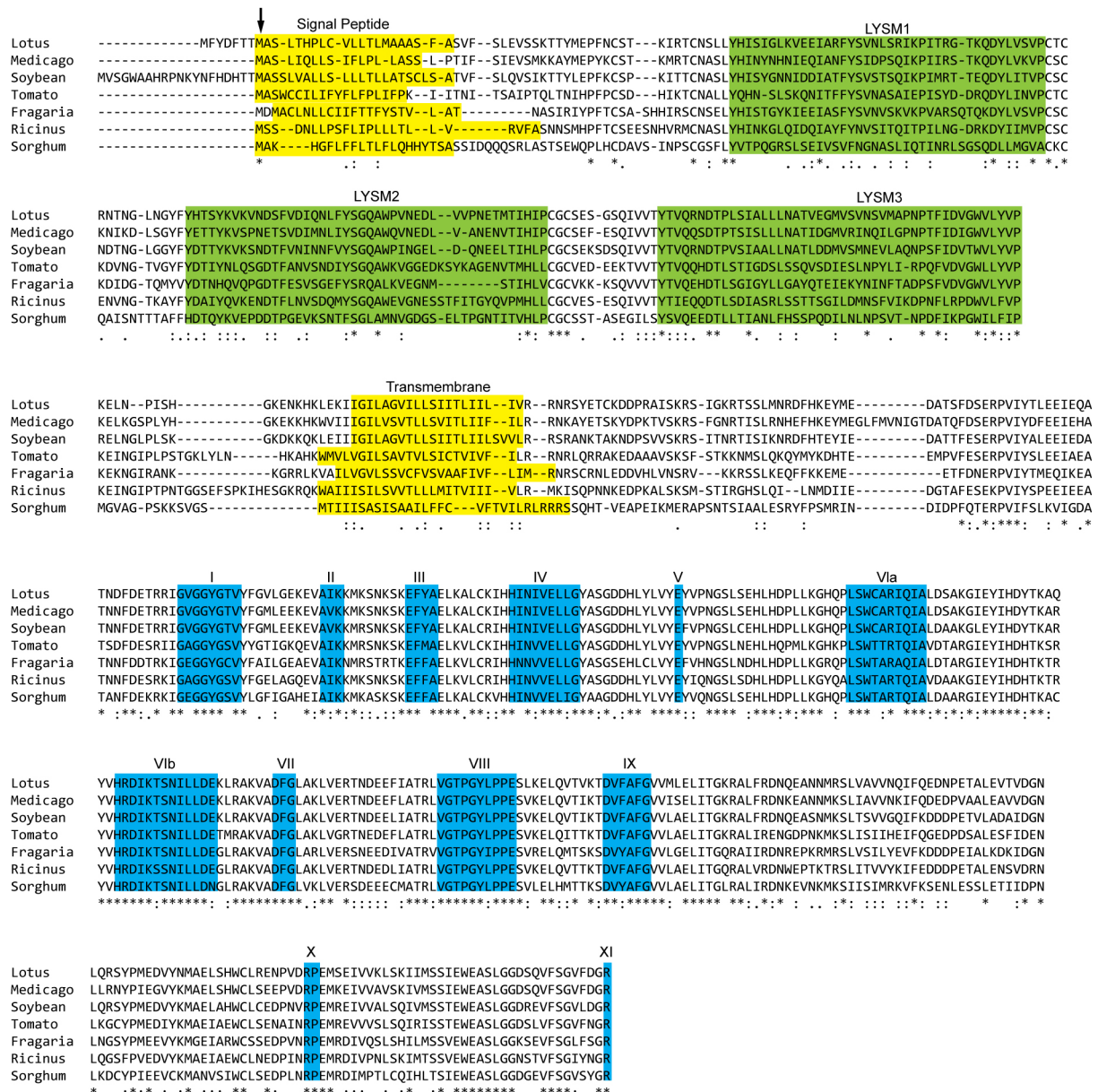
- Handberg, K. & Stougaard, J. *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J.* **2**, 487–496 (1992).
- Broughton, W. J. & Dilworth, M. J. Control of leghaemoglobin synthesis in snake beans. *Biochem. J.* **125**, 1075–1080 (1971).
- Sullivan, J. T. et al. Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. *J. Bacteriol.* **184**, 3086–3095 (2002).
- Kelly, S. et al. Genome sequence of the *Lotus* spp. microsymbiont *Mesorhizobium loti* strain R7A. *Stand. Genomic Sci.* **9**, 6 (2014).
- Leong, S. A., Williams, P. H. & Ditta, G. S. Analysis of the 5' regulatory region of the gene for delta-aminolevulinic acid synthetase of *Rhizobium meliloti*. *Nucleic Acids Res.* **13**, 5965–5976 (1985).
- Kawaharada, Y., Eda, S., Minamisawa, K. & Mitsui, H. A *Mesorhizobium loti* mutant with reduced glucan content shows defective invasion of its host plant *Lotus japonicus*. *Microbiology* **153**, 3983–3993 (2007).
- Hansen, J., Jørgensen, J.-E., Stougaard, J. & Marcker, K. Hairy roots — a short cut to transgenic root nodules. *Plant Cell Rep.* **8**, 12–15 (1989).
- Stougaard, J., Abildsten, D. & Marcker, K. The *Agrobacterium rhizogenes* pRi TL-DNA segment as a gene vector system for transformation of plants. *Mol. Gen. Genet.* **207**, 251–255 (1987).
- Petit, A. et al. Transformation and regeneration of the legume *Lotus corniculatus*: A system for molecular studies of symbiotic nitrogen fixation. *Mol. Gen. Genet.* **207**, 245–250 (1987).
- Thygesen, M. B., Sauer, J. & Jensen, K. J. Chemoselective capture of glycans for analysis on gold nanoparticles: carbohydrate oxime tautomers provide functional recognition by proteins. *Chemistry* **15**, 1649–1660 (2009).

51. Xiong, X. *et al.* Receptor binding by an H7N9 influenza virus from humans. *Nature* **499**, 496–499 (2013).
52. Heckmann, A. B. *et al.* Cytokinin induction of root nodule primordia in *Lotus japonicus* is regulated by a mechanism operating in the root cortex. *Mol. Plant Microbe Interact.* **24**, 1385–1395 (2011).
53. Czechowski, T., Stitt, M., Altmann, T., Udvardi, M. K. & Scheible, W. R. Genome-wide identification and testing of superior reference genes for transcript normalization in *Arabidopsis*. *Plant Physiol.* **139**, 5–17 (2005).
54. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**, 7 (2002).
55. Earley, K. W. *et al.* Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J.* **45**, 616–629 (2006).
56. Voinnet, O., Rivas, S., Mestre, P. & Baulcombe, D. An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* **33**, 949–956 (2003).
57. Hayashi, M. *et al.* Construction of a genetic linkage map of the model legume *Lotus japonicus* using an intraspecific F2 population. *DNA Res.* **8**, 301–310 (2001).
58. Sandal, N. *et al.* A genetic linkage map of the model legume *Lotus japonicus* and strategies for fast mapping of new loci. *Genetics* **161**, 1673–1683 (2002).
59. Krusell, L. *et al.* Shoot control of root development and nodulation is mediated by a receptor-like kinase. *Nature* **420**, 422–426 (2002).
60. Nishimura, R. *et al.* HAR1 mediates systemic regulation of symbiotic organ development. *Nature* **420**, 426–429 (2002).
61. Tirichine, L. *et al.* Deregulation of a Ca^{2+} /calmodulin-dependent kinase leads to spontaneous nodule development. *Nature* **441**, 1153–1156 (2006).
62. Tirichine, L. *et al.* A gain-of-function mutation in a cytokinin receptor triggers spontaneous root nodule organogenesis. *Science* **315**, 104–107 (2007).



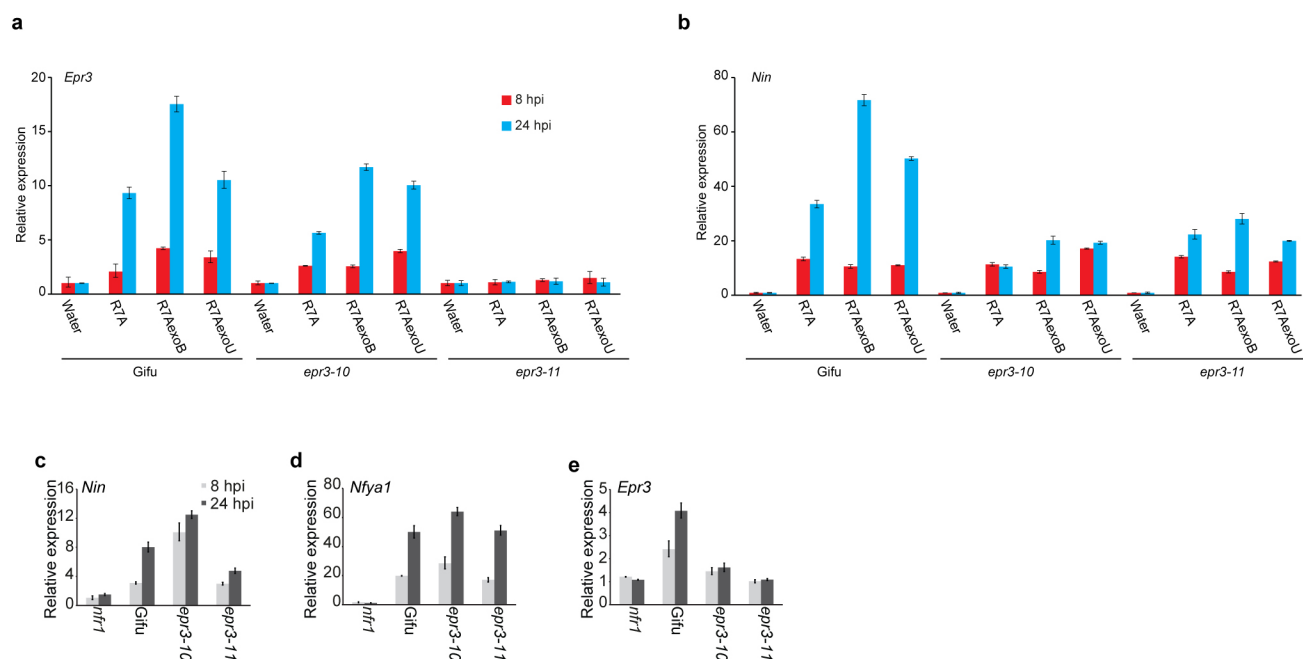
Extended Data Figure 1 | Genetic mapping and gene structure of *Epr3*.
a, Genetic map^{57,58} around the *Epr3* locus. Positions of bacterial artificial chromosome (BAC) and transformation-competent artificial chromosome (TAC) clones and of closest markers together with the number of informative

recombinations delimiting the *Epr3* locus are indicated. **b**, Alignment of genomic and cDNA sequences defined 10 exons in *Epr3* and a gene structure spanning 4,770 bp. The *Epr3* gene is marked together with the exon-intron structure of *Epr3* and positions of mutations.

a**b**

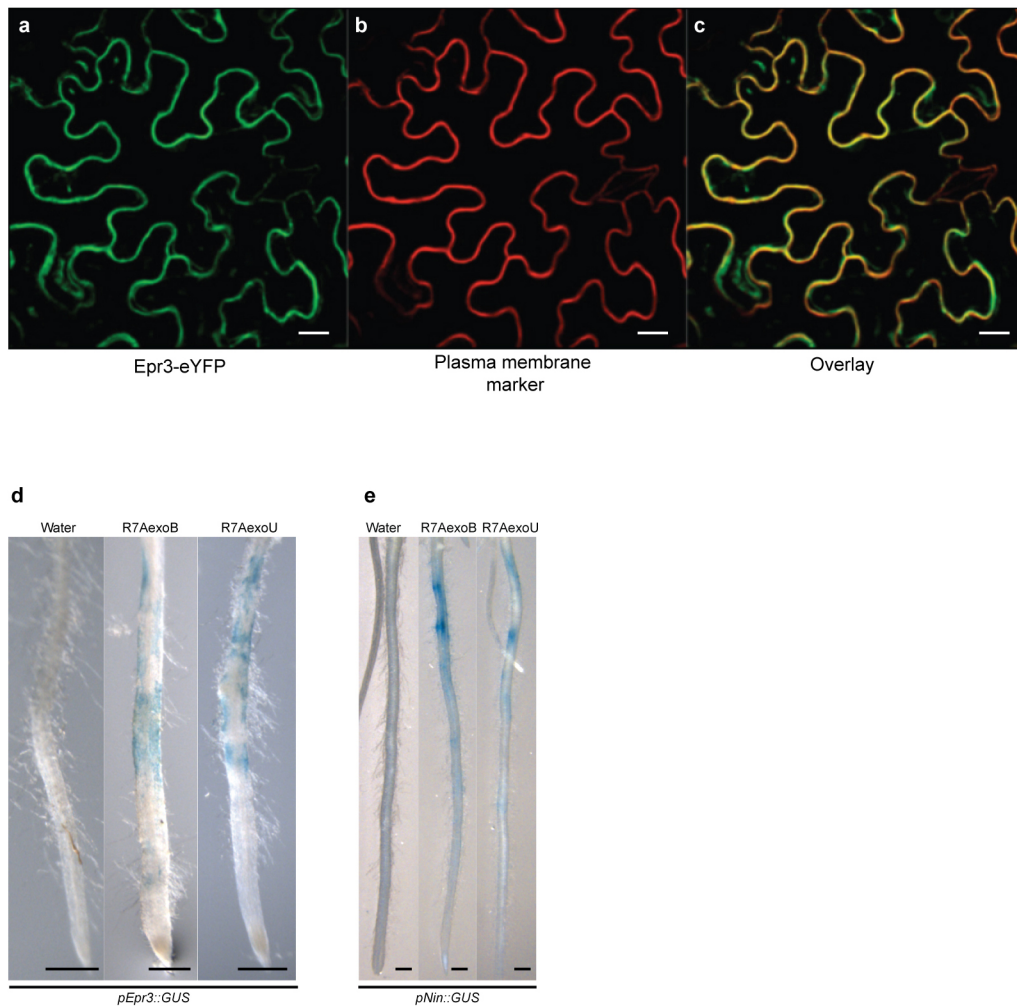
Extended Data Figure 2 | Domain structure of EPR3 and homologues proteins. **a**, Alignment of LysM modules and CxC motifs from EPR3 and plant homologues. The cysteine substituted in *epr3-3* and the conserved proline substituted in *epr3-10* are marked. Asterisks indicate conserved amino acids. **b**, Alignment of the full-length EPR3 protein and homologous full-length proteins from *Medicago* (XP_003613165.1), soybean (XP_003517716.1, XP_003530632.1), tomato (XP_004242179.1), *Ricinus* (XP_002527912.1), *Fragaria* (XP_004300916.1) and *Sorghum* (XP_002455766.1). Predicted

N-terminal signal peptides are marked and highlighted in yellow. For EPR3 a signal peptide of 21 amino acids is predicted by SignalP 4.1, assuming that the second of two methionines (arrow) that are eight amino acids apart is the correct N terminus. The three extracellular LysM domains are marked and highlighted in green. The putative transmembrane region is marked and highlighted in yellow. The classical kinase domains are marked I to XI and highlighted in blue.



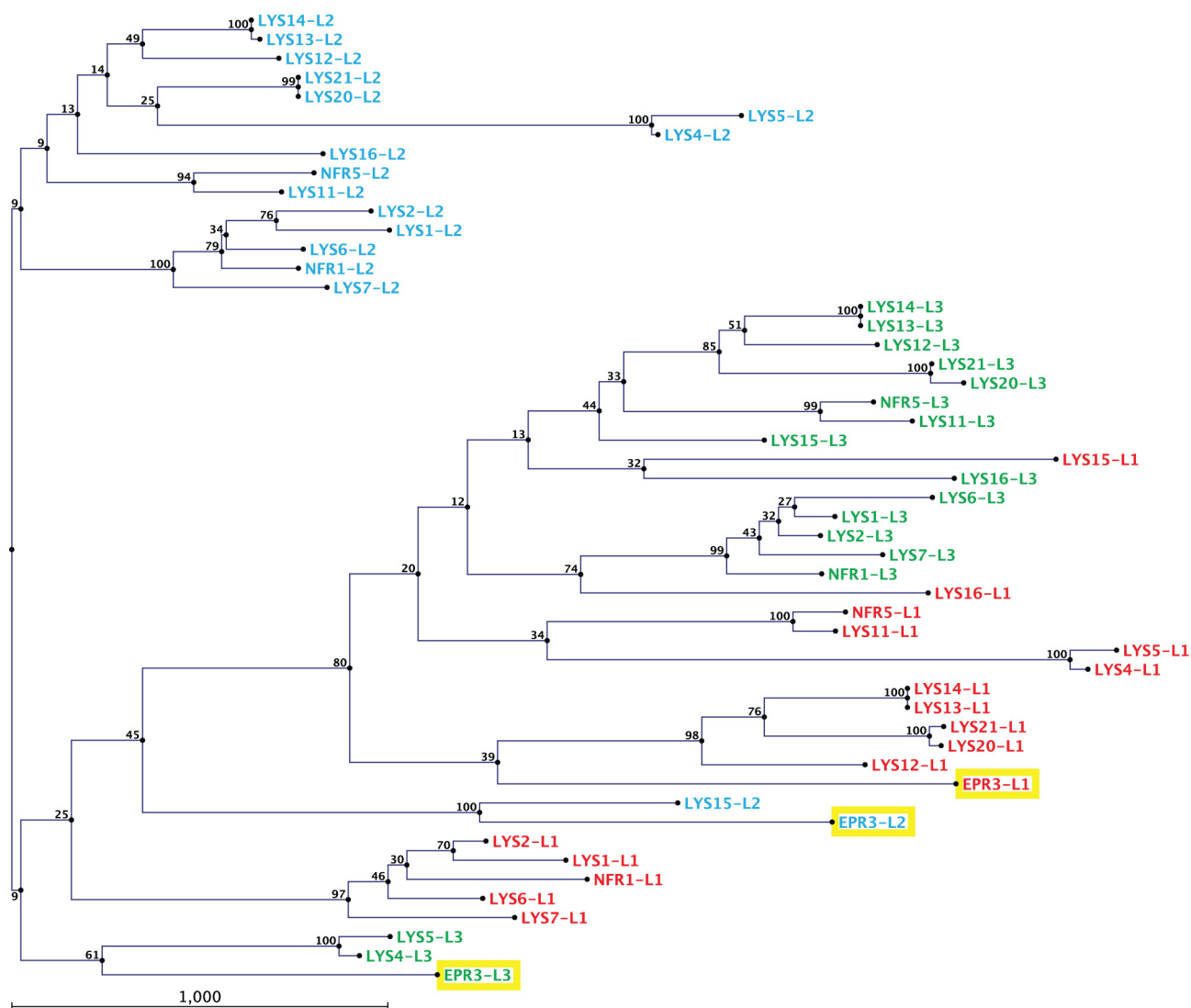
Extended Data Figure 3 | Quantification of transcript levels. **a, b**, Relative expression of *Epr3* (**a**) and *Nin* (**b**) transcript levels in Gifu, *epr3-10* and *epr3-11* mutants at 8 and 24 h post inoculation (h.p.i.) with R7A, R7AexoB and R7AexoU. **c–e**, Relative expression of *Nin*, *Nfya1* and *Epr3* transcripts in Gifu,

epr3-10 and *epr3-11* and *nfr1-1* roots treated with purified Nod factors. Values are from three biological (each consisting of ten plants) and three technical replicates. Bars show the corresponding upper and lower 95% confidence intervals.



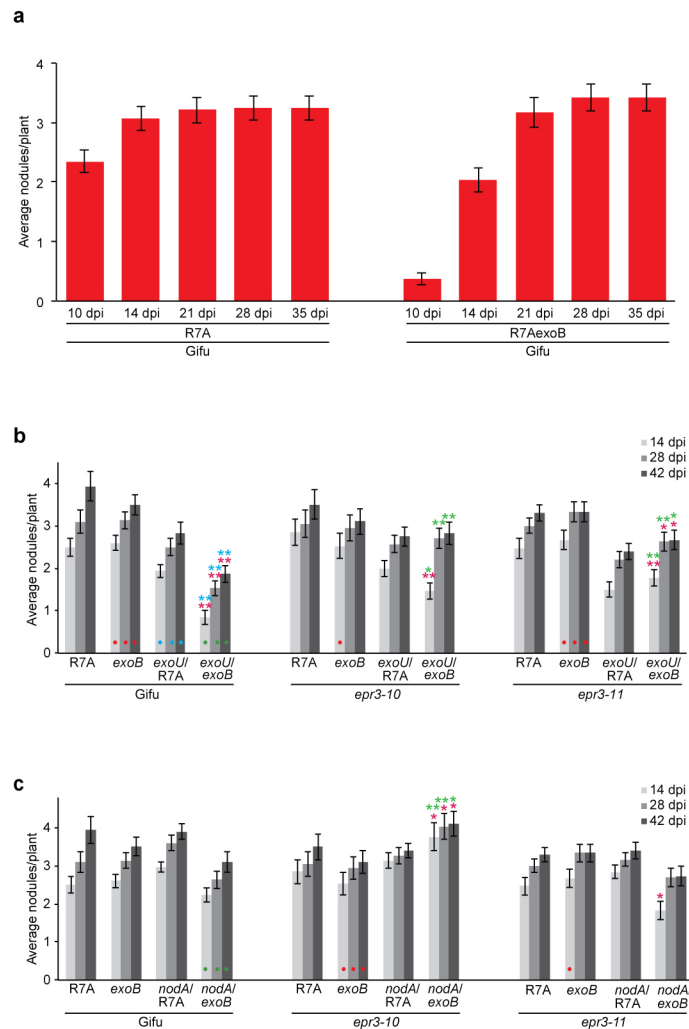
Extended Data Figure 4 | Localization of EPR3 protein and expression of *Epr3*. **a–c**, Membrane localization of EPR3. **a**, Representative image of EPR3–eYFP transiently expressed in *N. benthamiana* leaves shows co-localization with the plasma membrane dye FM 4-64 shown in **b**. An overlay is shown in

c. **d, e**, *pEpr3::GUS* (**d**) and *pNin::GUS* (**e**) expression in roots 14 and 4 days, respectively, after inoculation with R7AexoB and R7AexoU. In **d**, $n = 15$, $n = 26$ and $n = 10$ for water, R7AexoB and R7AexoU, respectively; in **e**, $n = 10$. Scale bars: **a–c**, 20 μm ; **d, e**, 0.5 mm.



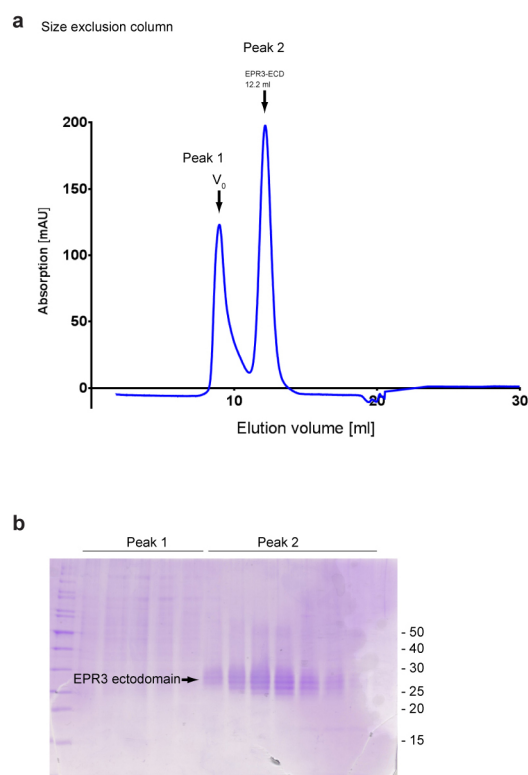
Extended Data Figure 5 | Phylogeny of the individual LysM modules of all seventeen *Lotus* LysM receptor kinases. Note that among these receptor kinases EPR3 stands out as the only protein where both the LysM2 and LysM3

modules do not cluster with the corresponding LysM2 and LysM3 modules of the other receptor kinases.



Extended Data Figure 6 | Competition for infection following R7AexoU pre-inoculation. **a**, Nodulation kinetics of Gifu inoculated with R7A ($n = 30$) or R7AexoB ($n = 29$). Nodule numbers represent infected nodules only. **b**, Time course of nodule numbers on Gifu, *epr3-10* and *epr3-11* inoculated with R7A and R7AexoB or pre-inoculated with R7AexoU followed by inoculation with R7A or R7AexoB 96 h later. $n = 30$, except for *epr3-10* R7A ($n = 20$), *epr3-10* R7AexoB ($n = 19$), *epr3-10* R7AexoU/R7A ($n = 28$) and *epr3-10* R7AexoU/

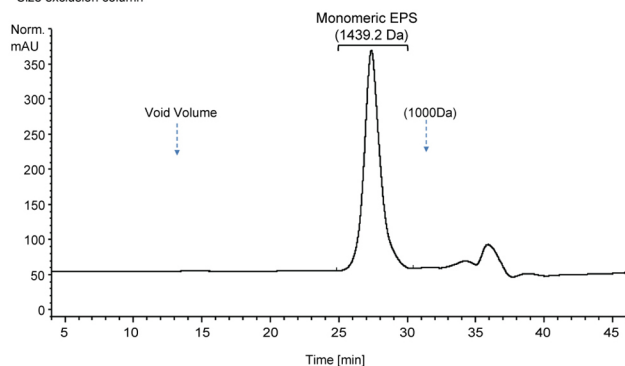
R7AexoB ($n = 28$). **c**, Control experiment. Time course of nodule numbers on Gifu, *epr3-10* and *epr3-11* inoculated with R7A and R7AexoB or pre-inoculated with R7A Δ nodA followed by inoculation with R7A or R7AexoB 96 h later. $n = 30$, except for *epr3-10* R7A ($n = 20$) and *epr3-10* R7A Δ nodA/R7AexoB ($n = 29$). $**P < 0.01$ and $*P < 0.05$ (t -test) indicate significant differences in nodule number compared to groups at the same d.p.i. and indicated by matching coloured dots. Error bars show s.e.m.



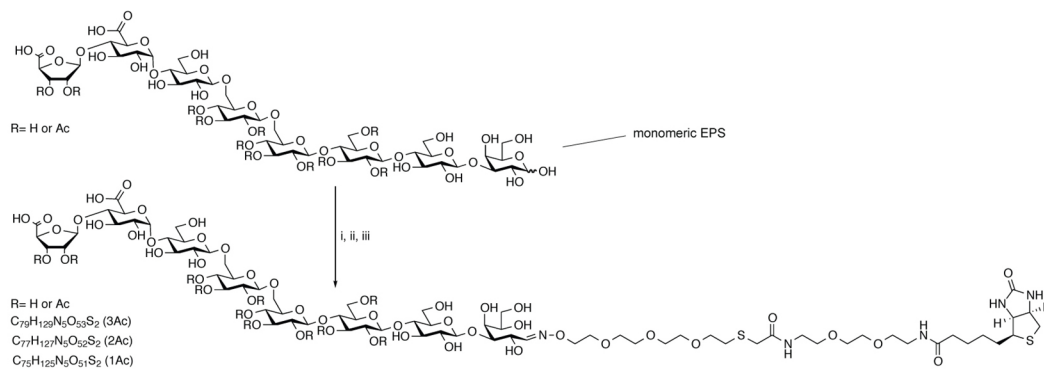
Extended Data Figure 7 | Purification of the EPR3 ectodomain. **a**, EPR3 ectodomain expressed in insect cells was purified twice using Ni-IMAC affinity columns followed by a Superdex 75 10/300 GL column size exclusion. In the SEC profile shown, pure EPR3 ectodomain elutes as a single peak (Peak 2), at 12.2 ml elution volume. Contaminants elute in the void volume (V_0 ; Peak 1).

mAU is UV absorbance at 280 nm. **b**, Corresponding 15% Coomassie blue-stained SDS-PAGE gel. The EPR3 ectodomain from Peak 2 in different glycosylation states is visible as distinct bands (black arrow). Units on the right hand side list the molecular weight (kilodaltons) of marker proteins in the corresponding position in the first lane to the left.

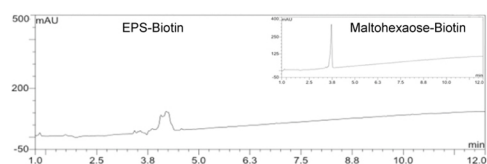
a Size exclusion column



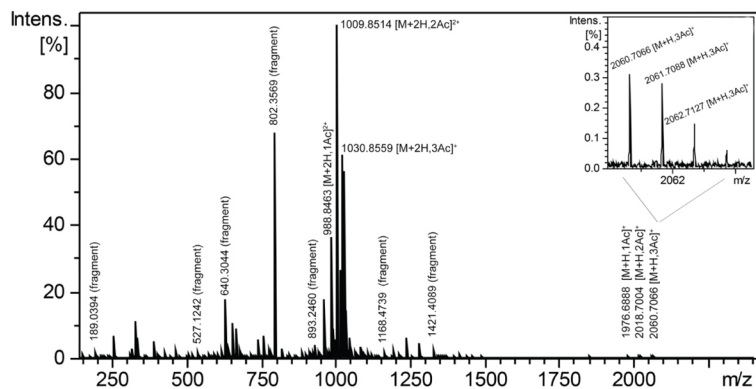
b



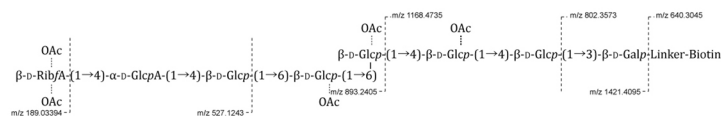
C



d



e



Extended Data Figure 8 | Isolation and characterization of EPS and the EPS–biotin conjugate by high-performance liquid chromatography and mass spectrometry. **a**, Size-exclusion chromatography profile of monomeric EPS isolated from culture media of R7AΔndvB. R7AΔndvB is deprived of a 2-linked cyclic glucan²⁴. The size-exclusion chromatography separation resolved one major fraction observed by refractive index detection, and the molecular size of the eluting fraction was assigned based on retention times of authentic standards of polysaccharides and MALDI–TOF mass spectrometry. The MALDI–TOF mass spectrometry analysis of the collected fraction was acquired in negative ionization mode and demonstrated the presence of a major monoisotopic $[M-H]^-$ ion at m/z 1,437.39 (calculated $[M-H]^- = 1,437.4055$, and formula weight = 1,439.19) consistent with a Hex₅GlcARibA octasaccharide substituted with three OAc groups. Vertical arrows mark

column void volume and retention time of a 1,000 Da standard. **b**, Synthesis of EPS octaose–biotin conjugate. Conditions: (i) Aminoxy-thiol OEG linker, aniline, acetate buffer, pH 4.5, 40 °C, 16 h; (ii) triethylsilane, trifluoroacetic acid, 15 min; (iii) biotin–iodoacetamide reagent, borate buffer, pH 8.3, 15 min. **c**, High-performance liquid chromatography chromatogram at 206 nm. The chromatogram displays several overlapping peaks, owing to a micro-heterogenous distribution of OAc groups, ranging from 1 to 3. The inset shows the chromatogram for the maltohexaose–biotin conjugate. **d**, Fourier transform ion cyclotron resonance (FT-ICR) mass spectrum of EPS octaose–biotin conjugate. Inset displays isotopic distribution of the $[M+H, 3Ac]^+$ ion. **e**, Structure and fragmentation pattern of EPS octaose–biotin conjugate. The displayed m/z values refer to **d**.

Extended Data Table 1 | The number of infected nodules and uninfected nodules ('bumps') on wild-type Gifu and *epr3* mutants 6 weeks after inoculation with R7AexoU *DsRed*

Plant genotype	No. of plants	Total no. of uninfected nodules	Total no. and % of infected nodules*	No. and % of plants forming infected nodules
Gifu	150	5909	4 (0.07%)	4 (2%)
<i>epr3-3</i>	60	1844	41 (2.2%)	27 (45%)
<i>epr3-9</i>	109	1982	31 (1.5%)	26 (24%)
<i>epr3-10</i> (exo277)	100	2805	61 (2.1%)	47 (47%)
<i>epr3-11</i>	181	5963	65 (1.1%)	58 (32%)
<i>epr3-12</i>	40	674	16 (2.3%)	12 (30%)
<i>epr3-13</i>	163	4409	54 (1.2%)	43 (26%)

* These numbers include both red and large white nodules infected with R7AexoU.

Extended Data Table 2 | Infection of plant mutants.

a

	Total plants	Uninfected nodules	Infected nodules
<i>snf1har1</i>	50	1321	0
<i>snf2har1</i>	44	342	1

b

Plant Genotype	Construct	No. of transformed plants	No. of transformants with nodules
Wild type	empty vector	29	29
<i>nfr1</i>	empty vector	30	0
<i>nfr1</i>	<i>p35S::Nfr1::t35S</i>	34	29
<i>nfr5</i>	<i>p35S::Nfr5::t35S</i>	31	26
Wild type	<i>p35S::Epr3::t35S</i>	34	33
<i>nfr1</i>	<i>p35S::Epr3::t35S</i>	9	0
<i>nfr5</i>	<i>p35S::Epr3::t35S</i>	22	0
<i>nfr1 nfr5</i>	<i>p35S::Epr3::t35S</i>	27	0
<i>nfr1</i>	<i>pNfr1::Nfr1</i>	52	47
<i>nfr1</i>	<i>pEpr3::Epr3</i>	35	0
<i>nfr1 nfr5</i>	<i>pNfr1::Nfr1 pNfr5::Nfr5</i>	10	6

a, The number of infected nodules on *snf1har1*⁵⁹⁻⁶¹ and *snf2har1*⁶² double mutants 6 weeks after inoculation with the R7AexoU DsRed strain. **b**, Complementation studies in *nfr1-1* or *nfr5-2* single mutants and *nfr1-1;nfr5-2* double mutants. *Epr3*, *Nfr1*, and *Nfr5* sequences were expressed from cognate promoters or the 35S promoter using a cassette with a 35S terminator.

Progesterone receptor modulates ER α action in breast cancer

Hisham Mohammed¹, I. Alasdair Russell¹, Rory Stark¹, Oscar M. Rueda¹, Theresa E. Hickey², Gerard A. Tarulli², Aurelien A. A. Serandour¹, Stephen N. Birrell², Alejandra Bruna¹, Amel Saadi¹, Suraj Menon¹, James Hadfield¹, Michelle Pugh¹, Ganesh V. Raj³, Gordon D. Brown¹, Clive D'Santos¹, Jessica L. L. Robinson¹, Grace Silva⁴, Rosalind Launchbury¹, Charles M. Perou⁴, John Stingl¹, Carlos Caldas^{1,5,6}, Wayne D. Tilley²§ & Jason S. Carroll¹§

Progesterone receptor (PR) expression is used as a biomarker of oestrogen receptor- α (ER α) function and breast cancer prognosis. Here we show that PR is not merely an ER α -induced gene target, but is also an ER α -associated protein that modulates its behaviour. In the presence of agonist ligands, PR associates with ER α to direct ER α chromatin binding events within breast cancer cells, resulting in a unique gene expression programme that is associated with good clinical outcome. Progesterone inhibited oestrogen-mediated growth of ER α ⁺ cell line xenografts and primary ER α ⁺ breast tumour explants, and had increased anti-proliferative effects when coupled with an ER α antagonist. Copy number loss of *PGR*, the gene coding for PR, is a common feature in ER α ⁺ breast cancers, explaining lower PR levels in a subset of cases. Our findings indicate that PR functions as a molecular rheostat to control ER α chromatin binding and transcriptional activity, which has important implications for prognosis and therapeutic interventions.

There is compelling evidence that inclusion of a progestogen as part of hormone replacement therapy increases the risk of breast cancer, implying that PR signalling can contribute towards tumour formation¹. However, the increased risk of breast cancer associated with progestogen-containing hormone replacement therapy is mainly attributed to specific synthetic progestins, in particular medroxyprogesterone acetate (MPA), which is known to also have androgenic properties². The relative risk is not significant when native progesterone is used³. In ER α ⁺ breast cancers, PR is often used as a positive prognostic marker of disease outcome⁴, but the functional role of PR signalling remains unclear. While activation of PR may promote breast cancer in some women and in some model systems, progesterone treatment has been shown to be antiproliferative in ER α ⁺ PR⁺ breast cancer cell lines^{5–7}, and progestogens have been shown to oppose oestrogen-stimulated growth of an ER α ⁺ PR⁺ patient-derived xenograft⁸. In addition, exogenous expression of PR in ER α ⁺ breast cancer cells blocks oestrogen-mediated proliferation and ER α transcriptional activity⁹. Furthermore, in ER α ⁺ breast cancer patients, PR is an independent predictor of response to adjuvant tamoxifen¹⁰, high levels of PR correlate with decreased metastatic events in early stage disease¹¹, and administration of a progesterone injection before surgery can provide improved clinical benefit¹². These observations imply that PR activation in the context of oestrogen-driven, ER α ⁺ breast cancer, can have an anti-tumorigenic effect. In support of this, PR agonists can exert clinical benefit in ER α ⁺ breast cancer patients that have relapsed on ER α antagonists¹³.

Breast cancers are typically assessed for ER α , PR and HER2 expression to define histological subtype and guide treatment options. PR is an ER α -induced gene¹⁴ and ER α ⁺ PR⁺ HER2[−] tumours tend to have the best clinical outcome because PR positivity is thought to reflect a tumour that is driven by an active ER α complex and therefore

likely to respond to endocrine agents such as tamoxifen or aromatase inhibitors^{10,15}. While ER α ⁺ PR⁺ tumours have a better clinical outcome than ER α ⁺ PR[−] tumours⁴, clinical response to ER α antagonists can vary, even among tumours with similar ER α and PR status^{15,16}, and recent evidence suggests that PR may be prognostic, but not predictive¹⁷. Some ER α ⁺ PR[−] tumours that are resistant to one class of ER α antagonists gain clinical benefit from another class, suggesting that in a subset of ER α ⁺ PR[−] breast cancers, the lack of PR expression does not reflect a non-functional ER α complex. It has been proposed that the non-functional ER α complex theory cannot completely explain PR negativity¹⁸. An alternative hypothesis is that other factors contribute to the loss of PR expression, which consequently influences breast tumour responses to ER α target therapies.

PR is recruited to the ER α complex

Given the controversial and complex interplay between the ER α and PR pathways in breast cancer, we explored the possible functional crosstalk between these two transcription factors and the implications for clinical prognosis in ER α ⁺ disease. Ligand-activated ER α and PR protein complexes were purified to ascertain interplay between these two transcription factors. Asynchronous ER α ⁺ PR⁺ MCF-7 and T-47D breast cancer cells were grown in stable isotope labelling by amino acids in cell culture (SILAC) growth media, which contains sufficient oestrogen to elicit maximal ER α binding to chromatin¹⁹. Oestrogen treatment is required to induce detectable levels of PR in MCF-7 cells, but not in T-47D cells²⁰. The two cell lines were subsequently treated with vehicle or one of two progestogens: native progesterone or the synthetic progestin R5020. Cells were cross-linked following hormone treatment and endogenous PR was immunopurified followed by mass spectrometry, using a technique we recently developed called RIME (rapid immunoprecipitation mass

¹Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, Cambridge CB2 0RE, UK. ²Dame Roma Mitchell Cancer Research Laboratories and the Adelaide Prostate Cancer Research Centre, School of Medicine, Hanson Institute Building, University of Adelaide, Adelaide, South Australia 5005, Australia. ³Department of Urology, University of Texas, Southwestern Medical Center at Dallas, Dallas, Texas 75390, USA. ⁴Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, 450 West Drive, CB7295, Chapel Hill, North Carolina 27599, USA. ⁵Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge University Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 2QQ, UK. ⁶Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK.

§These authors jointly supervised this work.

spectrometry of endogenous proteins)²¹. Under oestrogenic conditions, progesterone treatment induced an interaction between PR and ER α in the MCF-7 and T-47D cell lines, in support of previous findings showing a physical interaction between these two nuclear receptors²². In addition to ER α , progesterone treatment induced interactions between PR and known ER α -associated cofactors, including NRIP1, GATA3 and TLE3 (ref. 21) in both cell lines (Fig. 1a). As expected, treatment with natural ligand decreased interaction between PR and chaperone/co-chaperone proteins such as HSP90 and FKBP4/FKBP5 (Fig. 1a). The same findings were observed when R5020 was used as a synthetic ligand (Extended Data Fig. 1). Interestingly, when ER α was purified under the same treatment conditions, PR was the only differentially recruited protein in both cell lines (Fig. 1b). Moreover, the interaction between ER α and known ER α -associated cofactors was not differentially affected by progesterone treatment. A list of all interacting proteins under all experimental conditions is provided in Supplementary Table 1. The progesterone-induced ER α -PR interaction was confirmed by standard co-immunoprecipitation experiments in both MCF-7 and T-47D cells (Fig. 1c). We conclude that activation of PR results in a robust association between PR and the ER α complex. However, it remains unclear what effect this may have on ER α /PR DNA binding or what the primary DNA tethering mechanism may be (Fig. 1d).

Progesterone reprograms ER α binding events

Since PR is a transcription factor, we hypothesized that the progesterone-induced interaction between PR and ER α alters chromatin binding properties of the ER α -cofactor complex. MCF-7 and T-47D cell lines grown in complete (oestrogenic) media were stimulated with progesterone, R5020 or vehicle control for 3 or 24 h, followed by analysis of genome-wide ER α and PR and the co-activator p300 profiles by chromatin immunoprecipitation followed by sequencing (ChIP-seq). p300 deposits the H3K27Ac mark, which is indicative of functional enhancers²³. We found comparable ER α , PR and p300 binding at 3 h and 24 h (Extended Data Fig. 1 and data not shown) and chose 3 h for the remaining experiments. All ChIP-seq experiments were

subsequently repeated in triplicate following 3 h of treatment (sample clustering is provided in Extended Data Fig. 2). Whereas robust ER α binding (29,149 sites in MCF-7 cells and 8,438 sites in T-47D cells) was observed in oestrogenic conditions, limited PR binding events were seen under these conditions (Supplementary Table 2). Treatment with progesterone or the synthetic progestin R5020 in oestrogen-rich media resulted in robust PR recruitment to chromatin, with 46,191 PR binding events observed in T-47D cells with both ligands and 29,554 PR binding events in MCF-7 cells. Using DiffBind to identify differential peaks²⁴ that occurred following treatment of cells with progesterone or R5020, a rapid and robust redistribution of ER α binding to novel genomic loci was observed (Fig. 2a, b). In T-47D cells, 14,223 ER α binding events were reproducibly gained following 3 h of treatment with progesterone, a finding that was similar following stimulation with R5020, confirming a predictable redistribution of ER α chromatin binding following stimulation of the PR pathway. A similar rapid redirection of ER α was observed in MCF-7 cells after progesterone or R5020 treatment (Extended Data Fig. 1). The ER α sites reprogrammed by progesterone are likely to be functional, as indicated by the global recruitment of the co-activator p300 (Fig. 2b and Extended Data

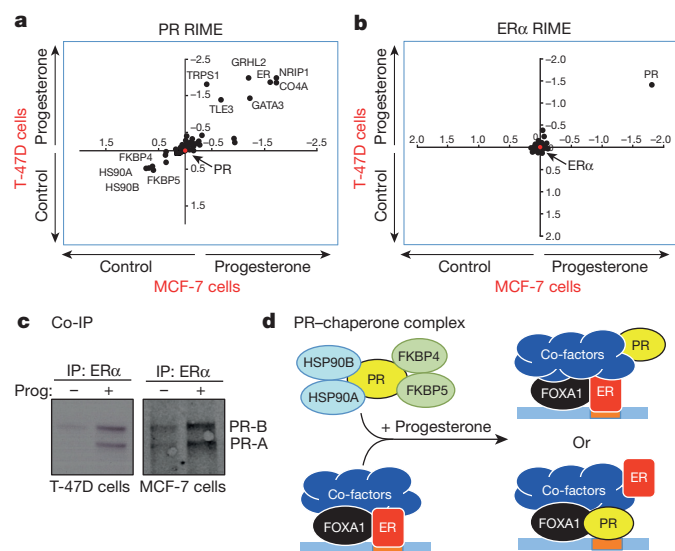


Figure 1 | PR is a novel ER α interacting protein following progesterone treatment. **a**, **b**, MCF-7 and T-47D breast cancer cells were SILAC-labelled, treated with hormones and harvested for RIME (endogenous protein purification-mass spectrometry) of either PR (**a**) or ER α (**b**). Differential proteins identified in both cell lines are plotted and all proteins are provided in Supplementary Table 1. The axes represent log fold change. **c**, Co-immunoprecipitation (IP) validation of PR interactions with ER α . Both PR isoforms interact with ER α . **d**, Model showing possible mechanisms of interplay between PR and ER α (ER)-cofactor complex.

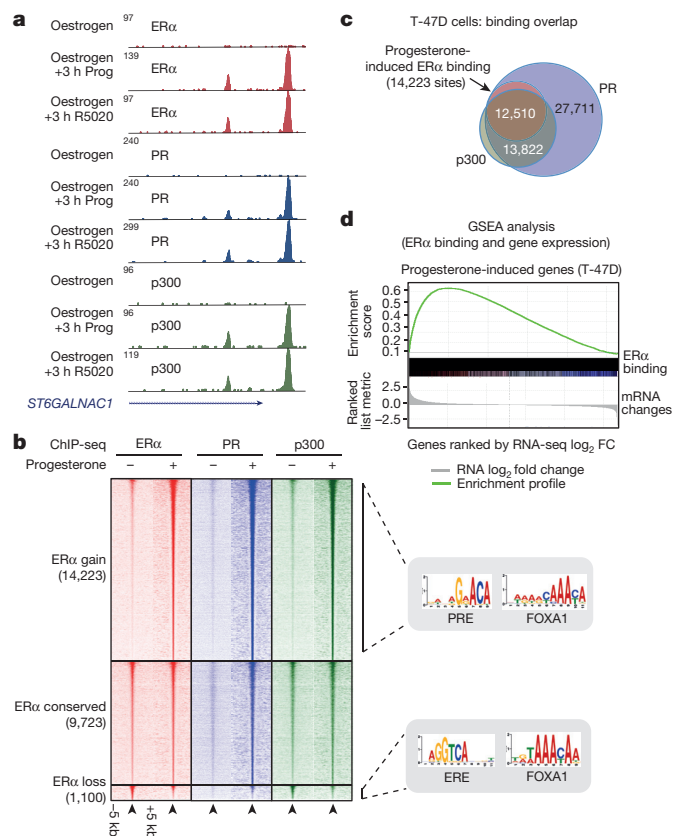


Figure 2 | Progesterone redirects oestrogen-stimulated ER α binding events to novel chromatin loci and transcriptional targets. ChIP-seq for ER α , PR and the co-activator p300 in T-47D cells grown in oestrogen-rich media and treated with progesterone or R5020. **a**, Example binding region of a progesterone/R5020 induced ER α , PR and p300 binding event. **b**, Heat map of ER α , PR and p300 ChIP-seq data from T-47D cells after 3 h of progesterone treatment. The heat map is shown in a horizontal window of ± 5 kb. Also shown are the enriched motifs within each category. ERE, oestrogen responsive elements; PRE, progesterone responsive element. **c**, Overlap between progesterone-induced ER α , PR and p300 binding sites, representing only the progesterone-induced ER α binding events. $n = 3$ independent biological replicates. **d**, RNA-seq was performed after progesterone or R5020 treatment for 3 h under oestrogenic conditions (control). $n = 8$ independent biological replicates. GSEA analysis was conducted, comparing progesterone-induced transcripts with progesterone-induced ER α binding events within T-47D cells.

Fig. 2). These ER α binding events appear to be mediated by PR, since 99% of the gained ER α binding sites in T-47D cells overlapped with a PR binding event (Fig. 2c) (the overlap in MCF-7 cells was 94%) and motif analysis of the gained ER α sites revealed the presence of progesterone responsive elements, but not oestrogen responsive elements (Fig. 2b). This suggests that PR mediates the interaction between the ER α -PR-p300 complex and DNA. ER α ChIP-seq was repeated in MCF-7 and T-47D cells following hormone deprivation and subsequent treatment with vehicle, oestrogen alone or progesterone alone. Single hormone treatments did not induce the ER α binding events observed under dual hormone conditions (Extended Data Fig. 3), confirming that the ER α reprogramming is dependent on having both receptors activated simultaneously. In addition, Forkhead motifs were enriched at the gained ER α sites and 49% of the gained ER α /PR binding events were shown to overlap with FOXA1 binding (Extended Data Fig. 4), consistent with previous findings showing that PR binding involves the pioneer factor FOXA1 (ref. 25). In keeping with the anti-proliferative effects of progestogens in ER α^+ breast cancer cells, analysis of the genes bound by ER α following progesterone treatment revealed enrichment for cell death, apoptosis and differentiation pathways (Supplementary Table 3).

To identify the transcriptional targets of the progesterone-induced ER α binding events, we treated oestrogen-stimulated MCF-7 and T-47D cells with progesterone, R5020 or vehicle for 3 h and performed eight replicates of RNA-seq. In total, 470 genes were differentially regulated by dual treatment with oestrogen and progesterone or R5020 compared to oestrogen alone in both T-47D and MCF7 cell lines (Extended Data Fig. 4). Gene set enrichment analysis (GSEA) analysis revealed a pronounced enrichment of progesterone-induced ER α binding events near genes upregulated by progesterone treatment in the presence of oestrogen (Fig. 2d). Collectively, these findings suggest that the progesterone-mediated changes in ER α binding events are functionally significant, since they co-recruit p300 and lead to new gene expression profiles (Fig. 2d and Extended Data Fig. 4). Importantly, increased expression of a gene signature that results from progesterone-stimulated ER α binding confers good prognosis in a cohort of 1,959 breast cancer patients (Extended Data Fig. 5).

The relative degree of ER α reprogramming and gene expression changes following progesterone treatment was higher in T-47D cells compared to MCF-7 cells (Supplementary Table 2), possibly owing to the differences in PR levels between these two cell lines²⁰. We assessed the PR gene (*PGR*) in these cell lines, which revealed copy number gain of the *PGR* gene in T-47D cells and a heterozygous loss of the *PGR* gene in MCF-7 cells (Extended Data Fig. 5). Exogenous expression of PR (both isoforms) in the MCF-7 cell line resulted in growth inhibition (Extended Data Fig. 6), confirming an anti-proliferative role for PR, via modulation of ER α transcriptional activity⁹.

Progesterone blocks ER α^+ tumour growth

To explore the hypothesis that progesterone stimulation could have beneficial effects on ER α^+ tumour growth *in vivo*, we established MCF-7-Luciferase xenografts in *NOD/SCID/IL2Rg^{-/-}* (NSG) mice and exposed the mice to control (that is, no hormone), slow release oestrogen pellets or slow release oestrogen plus standard high concentration progesterone pellets²⁶. Ten tumours for each condition were implanted (two tumours per mouse and five mice per condition) and tumour formation was monitored using bioluminescent imaging. After 25 days, in the absence of any hormone, tumours did not grow, but stimulation with oestrogen alone resulted in tumour growth. Co-treatment with oestrogen plus progesterone resulted in a significant decrease in tumour volume, as compared to oestrogen alone, when measured by bioluminescence (Fig. 3a, b) ($P = 0.0021$) or tumour volume ($P = 0.019$) (Extended Data Fig. 6). Immunohistochemistry confirmed that PR expression was induced under both oestrogen and oestrogen plus progesterone conditions (Extended Data Fig. 7), but importantly, PR will only be active under dual hormone conditions.

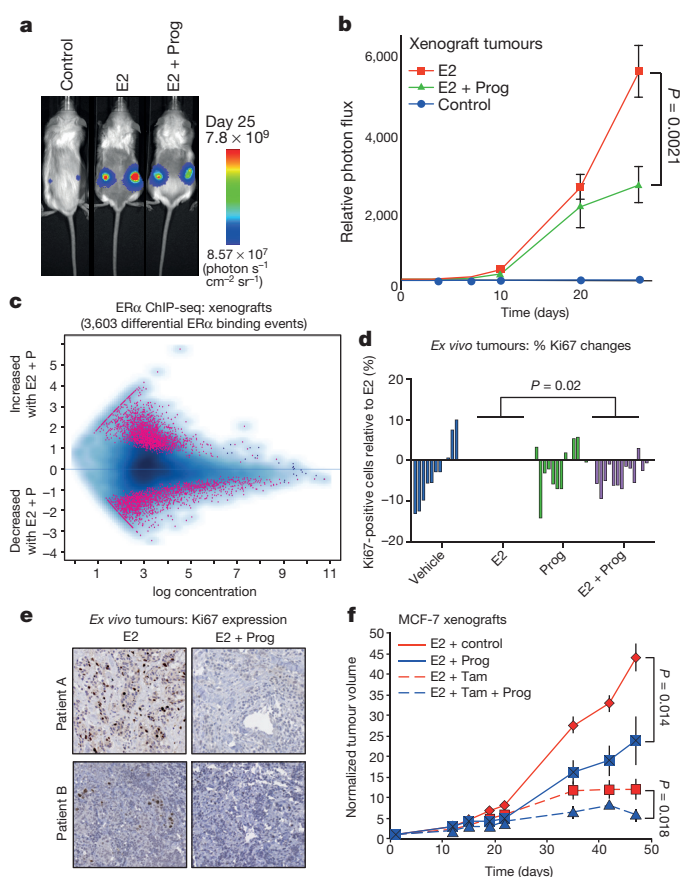


Figure 3 | Progesterone treatment inhibits ER α^+ tumour progression.

a, MCF-7-Luciferase cells were implanted in NSG mice with control, oestrogen (E2) pellets or oestrogen plus progesterone (E2 + Prog) pellets ($n = 10$). **b**, Graphical representation of tumour formation, as assessed by bioluminescence. **c**, ER α ChIP-seq was conducted on randomly chosen ($n = 6$) xenograft tumours from ovariectomized NSG mice treated with oestrogen alone or oestrogen plus progesterone. MA (log ratios and means average) plot representing changes in ER α binding. **d**, Proliferative responses (Ki67 staining) of primary breast cancer tissues cultured *ex vivo* with oestrogen (E2) or progestin (R5020) alone or both in combination ($n = 14$ samples per treatment; except for vehicle ($n = 11$) and R5020 treatments ($n = 12$)). The P value was calculated using a linear mixed effect analysis. **e**, Representative images of Ki67 immunostaining in *ex vivo* cultured breast tumour tissue sections from two patients. **f**, MCF-7 xenografts were grown in NSG mice in the presence of oestrogen pellets. Mice were treated with vehicle, tamoxifen, progesterone or tamoxifen plus progesterone and normalized tumour volume is shown. The data were analysed using a t -test and the error bars represent \pm s.e.m.

We repeated the MCF-7 xenograft experiment in ovariectomized mice, in order to eliminate any confounding issues related to endogenous mouse hormones. Assessment of tumour growth (Extended Data Fig. 7) confirmed the previous finding (Fig. 3b) that progesterone inhibited tumour formation. We performed ER α and PR ChIP-seq on six randomly selected sets of matched oestrogen or oestrogen plus progesterone stimulated xenograft tumours (taken from the final time point) and identified differentially bound ER α binding events. The major variable driving clustering of ER α binding events within tumour xenografts was the treatment condition (Extended Data Fig. 7). As observed in the short-term cell line experiments (Fig. 2), oestrogen-stimulated ER α binding in the xenograft tumours was substantially altered by progesterone treatment. We observed 3,603 differentially regulated ER α binding events in xenograft tumours from oestrogen plus progesterone conditions, when compared to oestrogen-only conditions (Fig. 3c). As such, progesterone induced a global reprogramming of ER α binding events *in vivo*, even following long-term hormonal treatment.

To extend our findings into primary tumours, we employed a novel *ex vivo* primary tumour culture system^{27,28} to cultivate ER α ⁺ PR⁺ primary tumours for short time periods, in order to study the effect of hormonal treatment on cell growth. Fourteen independent ER α ⁺ PR⁺ primary tumours were used for the analysis. Each tumour was cut into small pieces and randomized onto gelatine sponges half-submersed in media to sustain tissue architecture and viability. Tumour explants were then cultivated under hormone-deprived conditions for 36 h, followed by a 48 h treatment with vehicle control, oestrogen, the synthetic progestin R5020 or oestrogen plus R5020. Tumours retained normal cellular and morphological features after treatment in the *ex vivo* context (Extended Data Fig. 8). Fixed tissues were stained for Ki67 to assess changes in proliferation. Most of the tumours responded to oestrogen with a coincident increase in the percentage of cells that expressed Ki67. Co-treatment with progestin significantly inhibited oestrogen-stimulated proliferation (Fig. 3d and Extended Data Fig. 8), showing that progesterone can antagonise oestrogen-induced growth in primary human breast tumours cultivated as explants. Example images are shown in Fig. 3e. Importantly, the addition of progestin alone did not increase growth rates (Fig. 3d and Extended Data Fig. 8), in support of our hypothesis that progesterone lacks proliferative potential and importantly, is anti-proliferative in an oestrogen-driven context. ER α and PR co-localization was confirmed in the explant tumour samples using immunofluorescence, as was the reduction in Ki67 following progestogen treatment (Extended Data Fig. 8).

Given our findings that progestogens are anti-proliferative in an oestrogen-driven tumour context, we combined progesterone treatment with a standard of care therapy in xenograft models. MCF-7 xenografts were implanted in NSG mice and supplemented with slow-release oestrogen pellets. Mice were subsequently treated with vehicle, progesterone alone, tamoxifen alone or progesterone plus tamoxifen and tumour growth was monitored. As previously seen (Fig. 3b and Extended Data Fig. 6), progesterone antagonised oestrogen-induced tumour formation, as did tamoxifen alone, but the combination of tamoxifen plus progesterone had the greatest tumour inhibitory effect (Fig. 3f). This experiment was repeated in a second xenograft model (T-47D cells), confirming the finding that tumour volume was inhibited by progesterone alone, but the greatest degree of tumour inhibition was observed under conditions where an ER α antagonist (tamoxifen) was coupled with a PR agonist (progesterone) (Extended Data Fig. 6).

PGR copy number alterations are common

Given the *PGR* copy number loss observed in the MCF-7 cell line (Extended Data Fig. 5), we explored whether this was a common phenomenon observed in breast cancer patients, by assessing genomic copy number alterations (CNA) within the METABRIC cohort of ~2,000 breast cancers²⁹. This analysis revealed that 18.5% of all breast cancers possess a copy number loss in the *PGR* genomic locus and these are biased towards ER α ⁺ cases (Fig. 4a). This level of loss of heterozygosity (LOH) has been previously observed at this genomic locus in a small cohort of patients³⁰. In total, 21% of ER α ⁺ breast cancers contain a heterozygous or a homozygous deletion of the *PGR* genomic locus and these tumours had significantly ($P < 0.001$) lower PR messenger RNA levels (Fig. 4b). Importantly, within the 1,484 ER α ⁺ cases, tumours with copy number loss of the *PGR* genomic region had a poorer clinical outcome ($P = 0.001$) (Fig. 4c), suggesting that inactivation of the *PGR* gene contributes to worse outcome within this subset of ER α tumours.

ER α mRNA levels are significantly higher in tumours with *PGR* copy number loss (Fig. 4d). Copy number loss of *PGR* affects gene expression events, since the expression level of genes shown to be progestogen-induced under oestrogenic conditions in the cell lines was significantly lower ($P < 0.001$) in tumours with *PGR* copy number loss and, inversely, the expression level

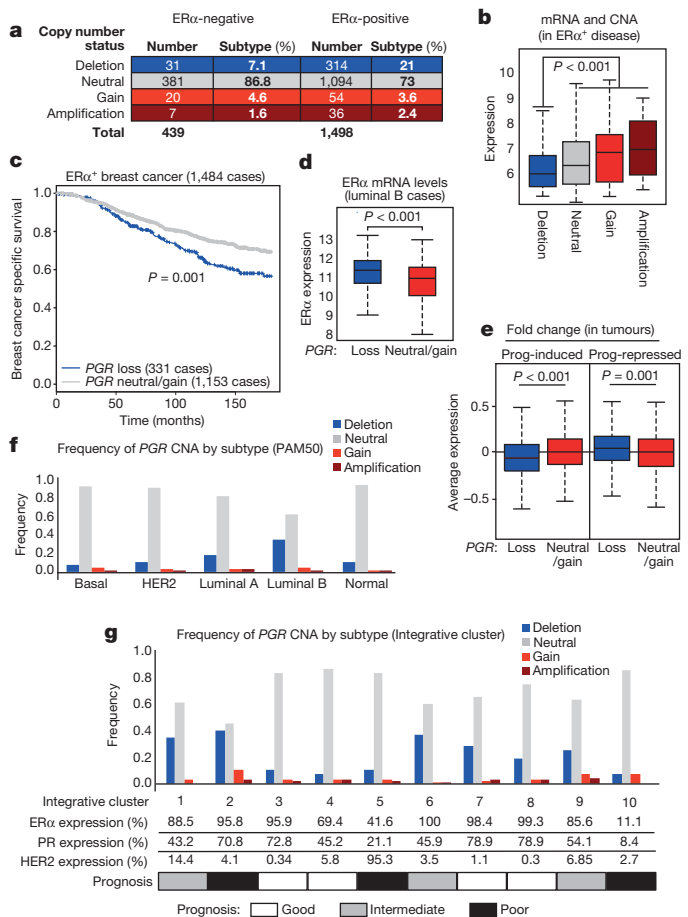


Figure 4 | The *PGR* genomic locus undergoes copy number loss in ER α ⁺ breast cancer. **a**, METABRIC breast cancers (1,937 in total) were assessed for copy number change in the *PGR* genomic locus. **b**, Correlation between PR mRNA levels and copy number status within all ER α ⁺ cases. The estimate of differences is 0.3551, 95% confidence interval: [0.244, 0.463]. **c**, Kaplan–Meier curve showing breast cancer specific survival in all ER α ⁺ cases. Cox model analysis: hazard ratio = 1.46 [1.156, 1.843]. **d**, Changes in ER α mRNA levels in luminal B tumours within the METABRIC cases. The estimate of differences is 0.4436, 95% confidence interval: [0.269, 0.618]. **e**, The expression levels of the stringent progestogen-induced or -repressed genes from the cell lines cultured under oestrogenic conditions were assessed in the ER α ⁺ METABRIC tumours. **f**, **g**, Relative fraction of tumours with *PGR* copy number alterations (can) events within molecular subtypes, based on the PAM50 gene expression profile (**f**) and the ten integrative clusters (**g**).

of genes repressed by progestogens in cell lines was significantly elevated ($P = 0.001$) in tumours with copy number loss of the *PGR* gene (Fig. 4e).

Using PAM50 breast cancer subtype stratification, 33% of luminal B tumours (typically poor prognosis and PR low/negative) possessed copy number loss of the *PGR* genomic region (Fig. 4f), explaining the lower PR protein levels in one third of this breast cancer subtype. In addition, even within the luminal A tumours only (which tend to have a better outcome), 19% have a copy number loss of *PGR* and this subset of luminal A tumours have a poorer clinical outcome (Extended Data Fig. 9). Tumours were then stratified based on the ten recently described genomic integrative clusters²⁹, revealing three ER α ⁺ subtypes (integrative clusters 1, 2 and 6) with at least 35% *PGR* copy number loss (Fig. 4g). All three of these integrative clusters have intermediate or poor clinical outcome (Fig. 4g). Data showing genomic copy number changes on the chromosome 11 arm that encompasses the *PGR* gene are shown in Extended Data Fig. 9. This level of *PGR* copy number loss was even higher in an independent cohort of breast cancer patients: 29% of luminal (ER α ⁺) breast cancers from

TCGA³¹ (Extended Data Fig. 10). Within luminal B tumours only, 39% of TCGA breast cancer cases had a *PGR* copy number loss.

Our data show that PR and ER α are functionally linked in breast cancer cells with a greater complexity than previously recognized. Good prognosis luminal A ER α ⁺ PR⁺ breast tumours, when exposed to both oestrogens and progestogens, have an ER α chromatin binding profile that is dictated by PR binding events. The functional significance of this steroid receptor crosstalk is regulation of a gene expression program associated with low tumorigenicity; hence, better disease outcome. Genomic alterations in the *PGR* genomic locus seem to be a relatively common mechanism for reduction of PR expression, which may consequently lead to altered ER α chromatin binding and target gene expression patterns that increase breast tumorigenicity and confers a poor clinical outcome. This ER α –PR crosstalk may be directly influenced by many variables, including the relative receptor levels and the hormonal milieu. The presence or absence of oestrogen may significantly alter the outcome of PR–ER α interactions, such that the anti-proliferative effects of PR activation may be limited to oestrogenic conditions. Our findings show that PR is not simply a marker of a functional ER α complex¹⁸. Rather, we propose that PR is a critical determinant of ER α function owing to crosstalk between PR and ER α . In this scenario, under oestrogenic conditions, an activated PR functions as a proliferative brake in ER α ⁺ breast tumours by re-directing ER α chromatin binding and altering the expression of target genes that induce a switch from a proliferative to a more differentiated state⁶.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 January; accepted 22 May 2015.

Published online 8 July 2015.

- Chlebowski, R. T. *et al.* Estrogen plus progestin and breast cancer incidence and mortality in postmenopausal women. *J. Am. Med. Assoc.* **304**, 1684–1692 (2010).
- Bentel, J. M. *et al.* Androgen receptor agonist activity of the synthetic progestin, medroxyprogesterone acetate, in human breast cancer cells. *Mol. Cell. Endocrinol.* **154**, 11–20 (1999).
- Fournier, A., Berrino, F. & Clavel-Chapelon, F. Unequal risks for breast cancer associated with different hormone replacement therapies: results from the E3N cohort study. *Breast Cancer Res. Treat.* **107**, 103–111 (2008).
- Blows, F. M. *et al.* Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10,159 cases from 12 studies. *PLoS Med.* **7**, e1000279 (2010).
- Vignon, F., Bardon, S., Chabos, D. & Rochefort, H. Antiestrogenic effect of R5020, a synthetic progestin in human breast cancer cells in culture. *J. Clin. Endocrinol. Metab.* **56**, 1124–1130 (1983).
- Musgrove, E. A., Swarbrick, A., Lee, C. S., Cornish, A. L. & Sutherland, R. L. Mechanisms of cyclin-dependent kinase inactivation by progestins. *Mol. Cell. Biol.* **18**, 1812–1825 (1998).
- Chen, C. C., Hardy, D. B. & Mendelson, C. R. Progesterone receptor inhibits proliferation of human breast cancer cells via induction of MAPK phosphatase 1 (MKP-1/DUSP1). *J. Biol. Chem.* **286**, 43091–43102 (2011).
- Kabos, P. *et al.* Patient-derived luminal breast cancer xenografts retain hormone receptor heterogeneity and help define unique estrogen-dependent gene signatures. *Breast Cancer Res. Treat.* **135**, 415–432 (2012).
- Zheng, Z. Y., Bay, B. H., Aw, S. E. & Lin, V. C. A novel antiestrogenic mechanism in progesterone receptor-transfected breast cancer cells. *J. Biol. Chem.* **280**, 17480–17487 (2005).
- Bardou, V. J., Arpino, G., Elledge, R. M., Osborne, C. K. & Clark, G. M. Progesterone receptor status significantly improves outcome prediction over estrogen receptor status alone for adjuvant endocrine therapy in two large breast cancer databases. *J. Clin. Oncol.* **21**, 1973–1979 (2003).
- Pichon, M. F., Pallud, C., Brunet, M. & Milgrom, E. Relationship of presence of progesterone receptors to prognosis in early breast cancer. *Cancer Res.* **40**, 3357–3360 (1980).
- Badwe, R. *et al.* Single-injection depot progesterone before surgery and survival in women with operable breast cancer: a randomized controlled trial. *J. Clin. Oncol.* **29**, 2845–2851 (2011).
- Bines, J. *et al.* Activity of megestrol acetate in postmenopausal women with advanced breast cancer after nonsteroidal aromatase inhibitor failure: a phase II trial. *Ann. Oncol.* **25**, 831–836 (2014).
- Lee, Y. J. & Gorski, J. Estrogen-induced transcription of the progesterone receptor gene does not parallel estrogen receptor occupancy. *Proc. Natl Acad. Sci. USA* **93**, 15180–15184 (1996).
- Dowsett, M. *et al.* Retrospective analysis of time to recurrence in the ATAC trial according to hormone receptor status: an hypothesis-generating study. *J. Clin. Oncol.* **23**, 7512–7517 (2005).
- The Breast International Group (BIG) 1-98 Collaborative Group. A comparison of letrozole and tamoxifen in postmenopausal women with early breast cancer. *N. Engl. J. Med.* **353**, 2747–2757 (2005).
- Viale, G. *et al.* Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1-98. *J. Clin. Oncol.* **25**, 3846–3852 (2007).
- Cui, X., Schiff, R., Arpino, G., Osborne, C. K. & Lee, A. V. Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. *J. Clin. Oncol.* **23**, 7721–7735 (2005).
- Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature Genet.* **43**, 27–33 (2011).
- Read, L. D., Snider, C. E., Miller, J. S., Greene, G. L. & Katzenellenbogen, B. S. Ligand-modulated regulation of progesterone receptor messenger ribonucleic acid and protein in human breast cancer cell lines. *Mol. Endocrinol.* **2**, 263–271 (1988).
- Mohammed, H. *et al.* Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep.* **3**, 342–349 (2013).
- Ballaré, C. *et al.* Two domains of the progesterone receptor interact with the estrogen receptor and are required for progesterone activation of the c-Src/Erk pathway in mammalian cells. *Mol. Cell. Biol.* **23**, 1994–2008 (2003).
- Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112 (2009).
- Stark, R. & Brown, G. D. DiffBind: differential binding analysis of ChIP-Seq peak data. *Bioconductor* <http://bioconductor.org/packages/release/bioc/html/DiffBind.html> (2011).
- Clarke, C. L. & Graham, J. D. Non-overlapping progesterone receptor cisomes contribute to cell-specific transcriptional outcomes. *PLoS ONE* **7**, e35859 (2012).
- Liang, Y., Besch-Williford, C., Brekken, R. A. & Hyder, S. M. Progesterin-dependent progression of human breast tumor xenografts: a novel model for evaluating antitumor therapeutics. *Cancer Res.* **67**, 9929–9936 (2007).
- Centenera, M. M. *et al.* Evidence for efficacy of new Hsp90 inhibitors revealed by *ex vivo* culture of human prostate tumors. *Clin. Cancer Res.* **18**, 3562–3570 (2012).
- Dean, J. L. *et al.* Therapeutic response to CDK4/6 inhibition in breast cancer defined by *ex vivo* analyses of human tumors. *Cell Cycle* **11**, 2756–2761 (2012).
- Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
- Tomlinson, I. P., Nicolai, H., Solomon, E. & Bodmer, W. F. The frequency and mechanism of loss of heterozygosity on chromosome 11q in breast cancer. *J. Pathol.* **180**, 38–43 (1996).
- Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors would like to thank S. Leigh-Brown, the staff in the genomic core facility, S. Halim, the proteomic core facility and the bioinformatic core facility at Cancer Research UK. We acknowledge S. Jindal for pathology review, N. Ryan for technical assistance and S. Edwards for statistical analysis with *ex vivo* culture. The MCF7-LucYFP cells were a kind gift from N. Benaich. We thank H. Gronemeyer for the PR-A and PR-B expressing vectors. We would like to acknowledge the support of the University of Cambridge, Cancer Research UK and Hutchison Whampoa Limited. Research reported in this manuscript was supported by the National Cancer Institute of the National Institutes of Health under award number 5P30CA142543 (to University of Texas Southwestern) and Department of Defense grants W81XWH-12-1-0288-03 (GVR). W.D.T. is supported by grants from the National Health and Medical Research Council of Australia (ID 1008349; ID 1084416) and Cancer Australia (ID 627229). T.E.H. held a Fellowship Award from the US Department of Defense Breast Cancer Research Program (BCRP; W81XWH-11-1-0592) and currently is supported by a Florey Fellowship from the Royal Adelaide Hospital Research Foundation. J.S.C. is supported by an ERC starting grant and an EMBO Young investigator award.

Author Contributions Experimental work was conducted by H.M., I.A.R., T.E.H., G.A.T., A.A.S., A.B., A.S., C.D., J.L.L.R., R.L. and G.S. Computational analysis was conducted by R.S., O.M.R., S.M. and G.D.B. Clinical samples, information and support was provided by S.N.B., G.V.R., C.M.P. and C.C. *In vivo* work was conducted by J.S. Genomic work was conducted by J.H. and M.P. All experiments were overseen by W.D.T. and J.S.C. The manuscript was written by H.M., I.A.R., T.E.H., W.D.T. and J.S.C. with help from the other authors.

Author Information All microarray and ChIP-seq data are deposited in GEO with the accession number GSE68359. All proteomic data are deposited with the PRIDE database with the accession number PXD002104. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.S.C. (jason.carroll@cruk.cam.ac.uk) or W.D.T. (wayne.tilley@adelaide.edu.au).

METHODS

Cell lines and SILAC labelling of cell lines. MCF-7 and T-47D human cell lines were obtained directly from ATCC and grown in DMEM, supplemented with 10% FBS. All cell lines were regularly genotyped using STR profiling using the Promega GenePrint 10 system. Cell lines were regularly tested for mycoplasma infection.

Rapid IP-mass spectrometry of endogenous protein (RIME). Rapid immunoprecipitation-mass spectrometry experiments were performed as previously described²¹. MCF-7 and T-47D cells were grown in R/K-deficient SILAC DMEM (PAA; E15-086), 10% dialysed serum (Sigma-Aldrich; F0392), and supplemented with 800 mM L-lysine ¹³C₆¹⁵N₂ hydrochloride and 482 mM L-arginine ¹³C₆¹⁵N₄ hydrochloride (Sigma-Aldrich) for 'heavy'-labelled media or 800 mM L-lysine ¹²C₆¹⁴N₂ hydrochloride and 482 mM L-arginine ¹²C₆¹⁴N₄ hydrochloride for 'light'-labelled media. Antibodies used were against ER α (Santa Cruz - sc-543, lot- A2213) and PR (Santa Cruz- sc-7208, lot H2312). 20 μ g of each antibody was used for each RIME experiment.

Each RIME experiment was performed by mixing 20 million cells from each label after respective drug treatments. Cells were treated with either progesterone (100 nM), R5020 (10 nM) or vehicle (ethanol). Two replicates of each experiment was performed and the results were validated by switching the SILAC labels. The RIME method, mass spectrometry and data analysis were performed as previously described²¹.

Western blot and co-immunoprecipitation. Immunoprecipitation experiments for ER α and PR were performed using Santa Cruz antibodies. Antibodies used were against ER α (sc-543) and PR (sc-7208). Western blots for ER α and PR were performed using Novocastra antibodies (ER α : NCL-L-ER6F11 and PR: NCL-L-PGR-AB).

ChIP-seq. ChIP-seq experiments were performed as described³². Antibodies used were against ER α (Santa Cruz - sc-543, lot- A2213), PR (Santa Cruz- sc-7208, lot H2312) and p300 (Santa Cruz sc-585, lot -E2412). 10 μ g of antibody was used for each experiment. Cells were treated with either progesterone (100 nM), R5020 (10 nM) or vehicle (ethanol) and experiment was performed in triplicates.

For xenograft experiments, ER α ChIP-seq experiments were performed from six sets of randomly chosen tumours under oestrogen only or progesterone and oestrogen treatment condition. Also included were two tumours from vehicle (control) conditions. Animals were randomly assigned to treatment groups using a random number generator at the beginning of the study.

ChIP-seq data analysis. Statistical tests and cut-offs were selected based on published recommendations²⁴.

RNA-seq. RNA-sequencing experiments were performed in MCF-7 and T-47D cells. Cells were treated with progesterone (100 nM) or R5020 (10 nM) for 3 h and RNA extracted. Experiment was performed in eight replicates for each cell line.

RNA-seq analysis. Single-end 40-bp reads generated on the Illumina HiSeq sequencer were aligned to the human genome version GRCh37.64 using TopHat v2.0.4 (ref. 33). Read counts were then obtained using HTSeq-count v0.5.3p9 (<http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html>). Read counts were then normalized and tested for differential gene expression using the DESeq³⁴ workflow. Multiple testing correction was applied using the Benjamini-Hochberg method. Genes were selected as differentially expressed such that false discovery rate (FDR) < 0.01.

GSEA analysis. Integration of the RNA-seq data and the ChIP-seq DBA results were carried out using gene set enrichment analysis (GSEA)³⁵ as follows. All genes assessed by RNA-seq were ranked and weighted by their mean log₂ fold change on progesterone treatments. Lists of genes that overlapped with regions showing significant differential binding on progesterone treatment were derived. These data were then analysed using the GSEA v2.0.13 GSEA Preranked tool.

Gene signature analysis. We derive a prognostic gene signature of consisting of progesterone regulated genes. We included genes that were determined by RNA-seq (see above) as differentially upregulated in T47-D cells when subjected to progesterone/progestin, as well as having a differentially bound ER binding site with increased binding affinity (determined using DiffBind) within 10 kb of the transcription start site. 38 genes met these criteria (see Extended Data Fig. 5). This signature was validated using the METABRIC expression data set (15), showing that patients in the highest 10% of expression of genes in the signature exhibit significantly better breast cancer specific survival than patients in the lowest 10% of expression ($P = 7.36 \times 10^{-4}$); Extended Data Fig. 5).

As it has been shown that many gene signatures have significant power to predict outcome in breast cancer³⁶, we subjected the signature to additional statistical tests using the Bioconductor package SigCheck³⁷. We used SigCheck to generate a null distribution of 1,000 signatures consisting of 38 genes selected at random, and computed their survival P values using the survival data for the 392 patients with high or low expression over the progesterone induced genes. The progesterone induced signature performed in the 99th percentile, with 1.2% of random

signatures demonstrating an equal or lesser P value (empirical P value = 0.012; Extended Data Fig. 5). Considering previously identified cancer signatures, of the 189 oncogenic signatures identified in MSigDB³⁸, 11 had equal or lesser P values, placing the progesterone regulation signature in the 94th percentile (empirical P value = 0.058). All the code for analysis is available upon request.

Explant tissue experiments

Clinical samples. Breast tumour samples and relevant clinical data were obtained from women undergoing surgery at the Burnside Private Hospital, Adelaide, South Australia, with informed, written consent. This study was approved by the University of Adelaide Human Research Ethics Committee (approval numbers: H-065-2005; H-169-2011). Patients enrolled in the University of Texas, Southwestern study provided written consent allowing the use of discarded surgical samples for research purposes according to an Institutional Board-approved protocol. Tumours were obtained from the University of Texas, Southwestern Tissue Repository under Institutional Review Board (IRB) (STU 032011-187).

Tissue collection and processing. Excised tissue samples were delivered to the laboratory on ice within one hour following surgery and washed in culture medium comprised of phenol red-free RPMI (SAFC Biosciences, Kansas, USA), 200 mM glutamine (SAFC Biosciences), 1 \times antibiotic-antimycotic (Sigma-Aldrich), 10 μ g/ml insulin (Sigma-Aldrich), and 10 μ g/ml hydrocortisone (Sigma-Aldrich). Breast tumour tissues were cultured *ex vivo* as previously described²⁸. In brief, tumour pieces (explants) were pre-incubated on gelatine sponges (3–4 per sponge) for 36 h in culture media containing 10% FCS, followed by treatment with hormones as indicated in culture media containing 10% steroid depleted FCS. The conditions used were: vehicle, oestradiol (10 nM), R5020 (10 nM) and the combination of oestradiol and R5020 (both at 10 nM) with hormone treatment conducted for 48 h. Representative pieces of tissue were fixed in 4% formalin in phosphate-buffered saline (PBS) at 4 °C overnight and subsequently processed into paraffin blocks. Sections (2 μ m) were stained with haematoxylin and eosin and examined by a pathologist to confirm and quantify the presence/proportion of tumour cells.

Immunohistochemistry. Primary antibodies to detect ER α (ID5 1:300; DAKO M7047, Glostrup, Denmark), PR (1:1,000, Leica NCL-PGR-AB, Wetzlar, Germany) or Ki67 (MIB1 1:400; DAKO M7240 Glostrup, Denmark) were used in conjunction with a 1:400 dilution of a biotinylated anti-mouse secondary antibody for 30 min (DAKO E0433, Glostrup, Denmark) followed by incubation with horseradish peroxidase-conjugated streptavidin (DAKO P0397, Glostrup, Denmark). Visualization of immunostaining was performed using 3,3'-diaminobenzidine (Sigma D9015), as previously described³⁸.

Image capture and quantification of Ki67 immunostaining. Slides were scanned at 40 \times magnification (NanoZoomer Microscopy System, Hamamatsu, Japan) and the digitized images edited to exclude non-glandular structures. A minimum of 10 high-resolution images from NanoZoomer files of all segments were recorded. Brown pixels (Ki67) and blue pixels (haematoxylin) were extracted from each image using Adobe Photoshop and the Colour Range tool and a fuzziness factor of 20. Extracted pixels were converted to greyscale format and a consistent threshold applied. Absolute numbers and proportion of Ki67-positive cells were determined using ImageJ software and the Analyse Particles tool with a circularity index of 0.2–1. An average of 14,000 cells were quantified per sample. Accuracy of automatic quantification was determined using an independent set of images that were assessed by an offsite independent observer who manually counted Ki67-positive and Ki67-negative cells in high-resolution images systematically sampled from an individual patient sample (containing >6,500 cells). Identical images were analysed by automatic quantification with consistency between manual and automatic counting of >95%. A linear mixed-effects model of percentage Ki67-positive cells versus type of treatment, adjusting for clustering on patient, was used to assess statistical significance.

Immunofluorescence. 5- μ m sections were cut and adhered to Superfost UltraPlus slides (Menzel-Glasser Braunschweig, Germany) and baked on a warm block at 60 °C for 1 h. Sections were dewaxed in 3 \times 5 min immersions of xylene followed by 3 \times 5 min immersions in 100% ethanol. Sections were rehydrated in 2 \times 5 min immersions in 70% ethanol followed by 2 \times 5 min immersions in distilled H₂O. Next, antigen retrieval was performed in 600 ml of 1 mM Na-EDTA (pH 8.0) by heating in an 800 W microwave until boiling (approximately 5 min), allowing slides to stand for 5 min before microwaving at 50% power for an additional 5 min. Slides were allowed to cool in antigen retrieval buffer for 60 min before transferring to PBS for 5 min. Sections were encircled with a wax pen and primary antibody diluted in PBS with 10% normal goat serum was applied overnight at 4 °C (rat anti-ER α 1:100 (Abcam ab46186); mouse anti-PR 1:500 (Dako NCL-PGR-AB); rabbit anti-Ki67 1:100 (Abcam ab16667). Sections were washed twice for 5 min in PBS, followed by incubation with secondary antibody diluted (all at 1:400) in PBS with 10% normal goat serum for 30 min at room temperature (goat anti-rat Alexa488 (A11006); goat anti-mouse Alexa569 (A11031); goat

anti-rabbit Alexa647, (A21246.), Life Technologies). Sections were washed twice for 5 min in PBS, followed by incubation with 1 nM DAPI diluted in PBS for 2 min at room temperature. Sections were mounted under DAKO fluorescent mounting media (S3023) and each fluorescent channel captured separately using a Zeiss LSM700 confocal microscope.

PGR copy number analysis. Matched DNA and RNA were extracted for 1,980 tumours. The copy number analysis was performed using the Affymetrix SNP 6.0 platform. The arrays were first pre-processed and normalized using CRMAv2³⁹ method from the *aroma.affymetrix* R package. For each array, allelic-crosstalk calibration, probe sequence effects normalization, probe-level summarization and PCR fragment length normalization were performed. Then the intensities were normalized against a pool of 473 normals for those samples that had no matched normal or against their matched normal when available. The log-ratios were segmented using the CBS algorithm⁴⁰ in the *DNACopy* Bioconductor package. Then, callings into five groups (homozygous deletion, heterozygous deletion, neutral copy number, gain and amplification), were made using thresholds based on the variability of each sample and their proportion of normal contamination. Then, samples were classified as PGR loss if they showed any type of loss in any part of the gene.

RNA analysis was performed using Illumina HT-12 v3 platform and analysed using beadarray package⁴¹. BASH⁴² algorithm was employed to correct for spatial artefacts. Bead-level data were summarized and a selection of suitable probes based on their quality was done using the re-annotation of the Illumina HT-12v3 platform⁴³. The samples were classified into the five breast cancer subtypes using PAM50⁴⁴ and the 10 integrative clusters²⁹. Two-sided *t*-tests were performed between expression values and loss of PGR to determine significance. Kaplan–Meier estimates and log-rank tests were obtained using the survival R package (<http://CRAN.R-project.org/package=survival>). Comparison of the expression of the stringent progesterone/R5020 induced or repressed genes in samples with PGR lost and not lost was performed averaging the expression of all the genes and running a two-sided *t*-test.

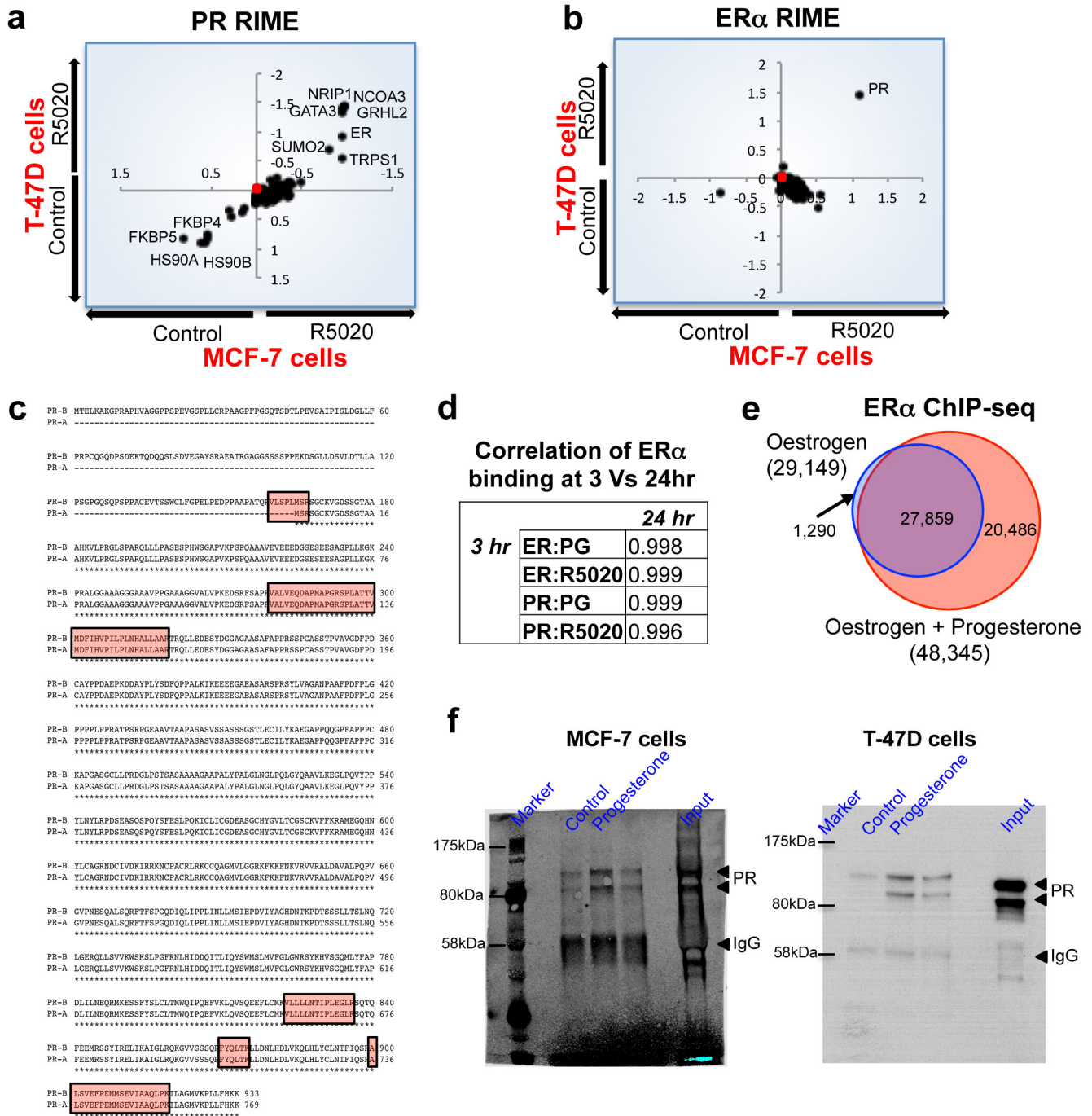
Xenograft experiments. Mice were maintained, and regulated procedures performed, according to UK Home Office project license guidelines. Sample size was calculated using a combination of the NC3Rs recommended Resource Equation method, and also prior knowledge of the experimental variability of breast cancer cell line xenografts in NSG animals. All *in vivo* experiments were performed using age matched female *NOD/SCID/IL2R^γ−/−* (NSG) mice and all experiments were blinded. Animals were selected using the following criteria: sex (female) and age matched (where possible littermates were used) to reduce variability; only animals in excellent health were used (verified pathogen free and in excellent physical health); all animals were certified as NSG strain. Animals were randomly assigned to treatment groups using a random number generator at the beginning of the study. Briefly, mice were injected subcutaneously into the No. 4 inguinal mammary fat pad with a suspension of 10⁵ MCF7-Luc2/YFP cells in 50% growth factor reduced Matrigel (BD Biosciences). Where appropriate, 90 day slow-releasing 17 β -oestradiol (0.72 mg per pellet) and/or progesterone (10 mg per pellet) hormone pellets (Innovative Research of America) were implanted subcutaneously in recipient mice. Tumour growth was monitored using regular bioluminescent imaging (IVIS) and caliper measurement. Data were analysed using the GraphPad Prism statistical software package. Standard deviations within groups were similar and substantially less than the variation between the treatment groups. Sample size was determined by a pilot experiment with a

smaller number of mice. Mice used for all experiments were age and litter matched (3 months) to reduce variability. 15 mice were injected with tumours and then 5 mice were selected at random for each treatment arm. Treatments were blinded using coded cages. Mice were regularly assessed for health and endpoint was determined on regulatory guidelines for tumour size.

For tamoxifen experiments, age-matched female NSG mice were injected as above with either 10⁵ MCF7-Luc2/YFP cells or 10⁷ T-47D-Luc2/mStrawberry cells in 50% growth factor reduced Matrigel (BD Biosciences). Concomitantly, a 90 day slow-release 17 β -oestradiol (0.72 mg per pellet) (Innovative Research of America) was implanted subcutaneously. One week later either 90 day slow-release progesterone (10 mg per pellet) hormone or placebo pellets (Innovative Research of America) were implanted subcutaneously, and tamoxifen/vehicle administration was commenced. 100 μ l tamoxifen (5 mg/mL) or vehicle (sterile filtered corn oil) were administered i.p. to a schedule of 3 days dosing/1 day rest for a total duration of 4 weeks (MCF7) or 8 weeks (T-47D). Tumour sizes were monitored as above. Each treatment arm consisted of 10 tumours.

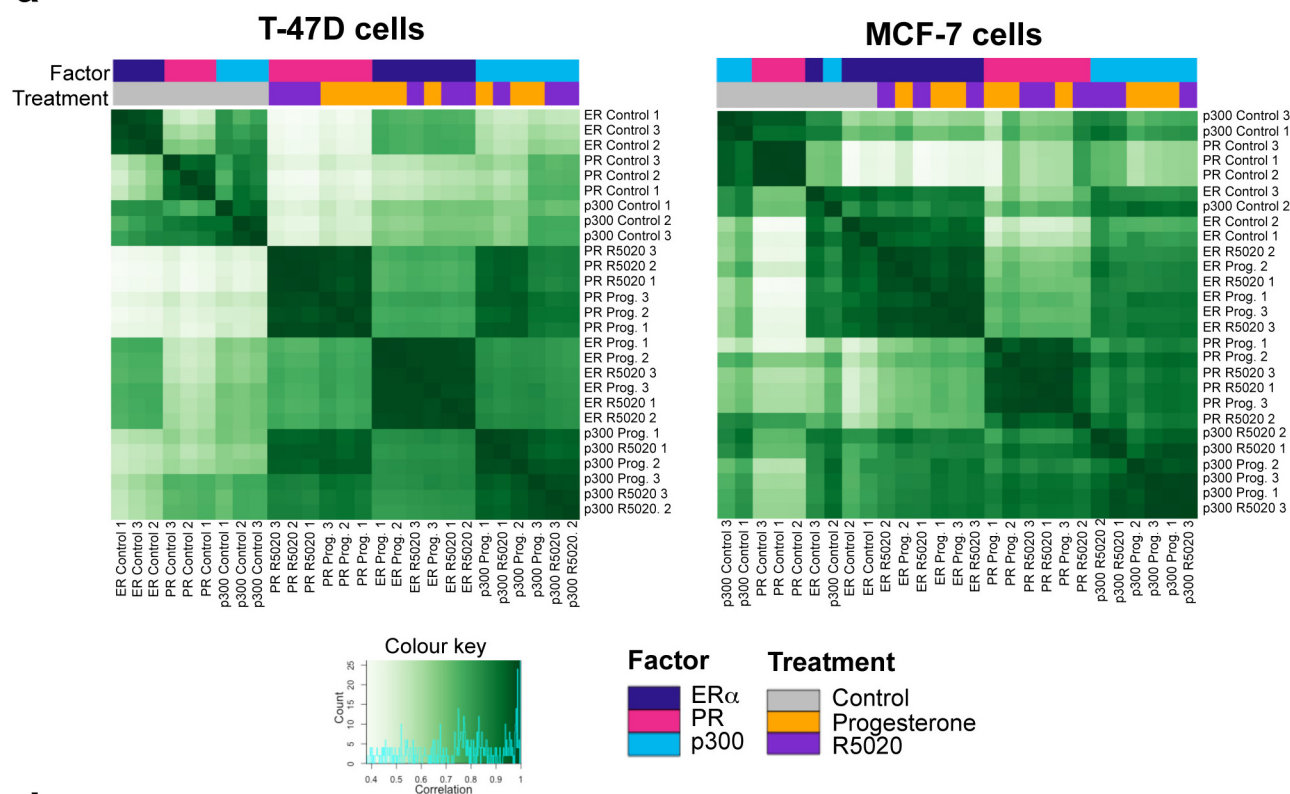
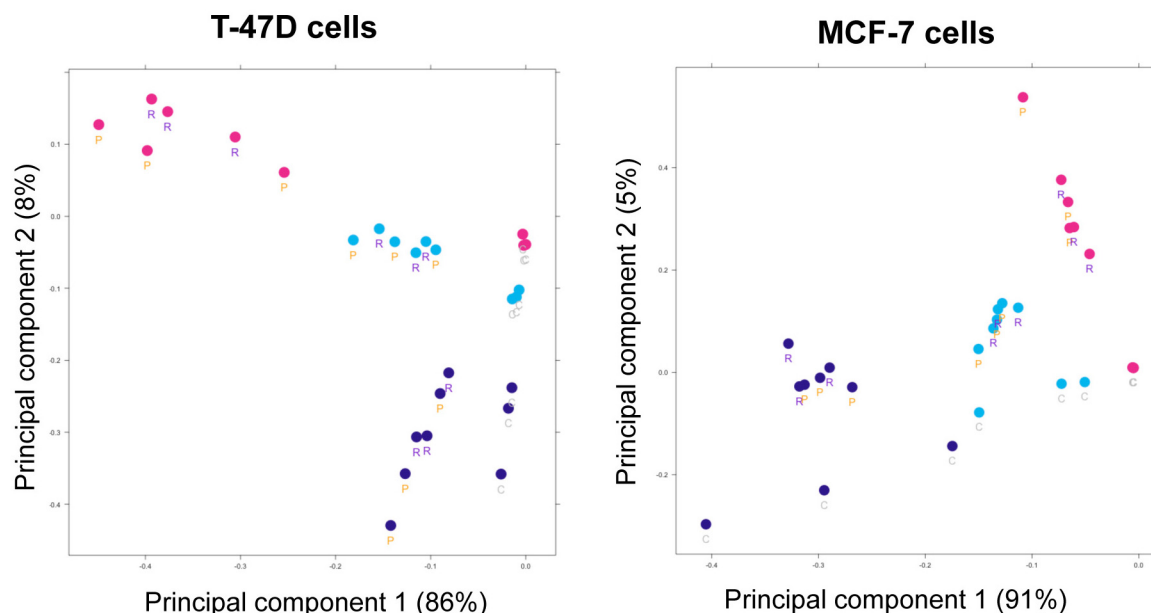
For ovariectomy experiments, age matched female NSG mice were injected as above with 10⁵ MCF7-Luc2/YFP cells in 50% growth factor reduced Matrigel (BD Biosciences), and 90 day slow-release 17 β -oestradiol (0.72 mg per pellet) and/or progesterone (10 mg per pellet) and/or placebo hormone pellets (Innovative Research of America) were implanted subcutaneously in recipient mice. Concomitantly, ovariectomy was performed. Tumour sizes were monitored as above for 7 weeks. Each treatment arm consisted of 10 tumours.

32. Schmidt, D. *et al.* ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**, 240–248 (2009).
33. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
34. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
35. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
36. Venet, D., Dumont, J. E. & Detours, V. Most random gene expression signatures are significantly associated with breast cancer outcome. *PLOS Comput. Biol.* **7**, e1002240 (2011).
37. Stark, R. & Norden, J. Checking gene expression signatures against random and known signatures with SigCheck. *Bioconductor* <http://www.bioconductor.org/packages/release/bioc/html/SigCheck.html>.
38. Tilley, W. D. *et al.* Detection of discrete androgen receptor epitopes in prostate cancer by immunostaining: measurement by color video image analysis. *Cancer Res.* **54**, 4096–4102 (1994).
39. Bengtsson, H., Wirapati, P. & Speed, T. P. A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6. *Bioinformatics* **25**, 2149–2156 (2009).
40. Venkatraman, E. S. & Olshen, A. B. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* **23**, 657–663 (2007).
41. Dunning, M. J., Smith, M. L., Ritchie, M. E. & Tavare, S. beadarray: R classes and methods for Illumina bead-based data. *Bioinformatics* **23**, 2183–2184 (2007).
42. Cairns, J. M., Dunning, M. J., Ritchie, M. E., Russell, R. & Lynch, A. G. BASH: a tool for managing BeadArray spatial artefacts. *Bioinformatics* **24**, 2921–2922 (2008).
43. Barbosa-Morais, N. L. *et al.* A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res.* **38**, e17 (2010).
44. Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).



Extended Data Figure 1 | Protein purification of ER α and PR interacting proteins, using RIME, following treatment with a synthetic progesterone. T-47D and MCF-7 breast cancer cells were grown in SILAC-isotope containing media and treated with either vehicle control or R5020, a synthetic progestin for 3 h. PR (a) or ER α (b) RIME was conducted and the proteins that were quantitatively enriched in both cell lines are shown. Only proteins that were enriched with a FDR < 1% were included. c, Peptide coverage of the PR protein following ER α RIME in T-47D cells. The identified peptides are highlighted and one of the peptides covers the 'Bus' region representing the PR-B isoform.

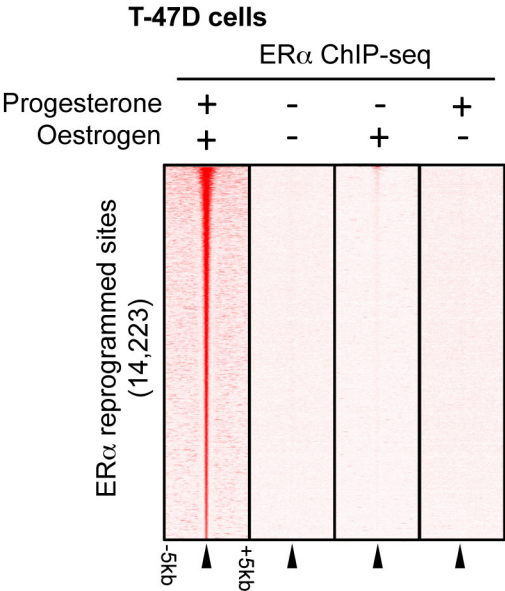
d, Comparison of binding at different time points and treatment of MCF-7 breast cancer cells with progesterone. ER α ChIP-seq at 3 h and 24 h results in comparable binding. Correlation between progesterone (PG) and R5020 (RO) at 3 and 24 h. e, MCF-7 cells were grown in oestrogen rich complete media and treated with progesterone or vehicle control for 3 h. ER α ChIP was conducted and peaks that occurred in at least two of three independent replicates were considered. Venn diagram showing the changes in ER α binding following progesterone treatment of MCF-7 cells. f, Uncropped western blots from Fig. 1c.

a**b**

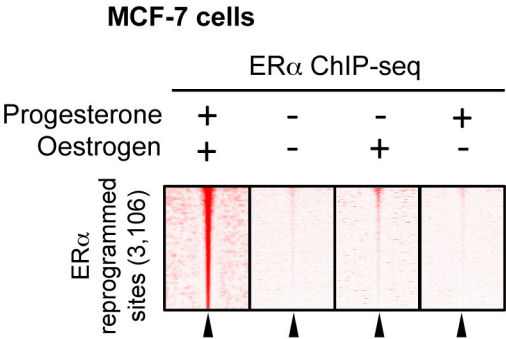
Extended Data Figure 2 | Clustering of ERα, PR, and p300 ChIP-seq experiments in two ERα⁺ cell lines. For each experiment, all binding sites identified as overlapping in at least two samples are merged and retained, and normalized read counts computed at each site for each sample. **a**, Clustering correlation heat maps, based on Pearson correlations read scores (with replicate

numbers in the labels), show good reproducibility between replicates and similarity of natural and synthetic hormone treatments. **b**, PCA plots showing the two most significant principal components (with samples labelled with treatment type: 'C' for full-media control conditions, 'P' for progesterone, and 'R' for R5020). The data from the two cell lines is shown.

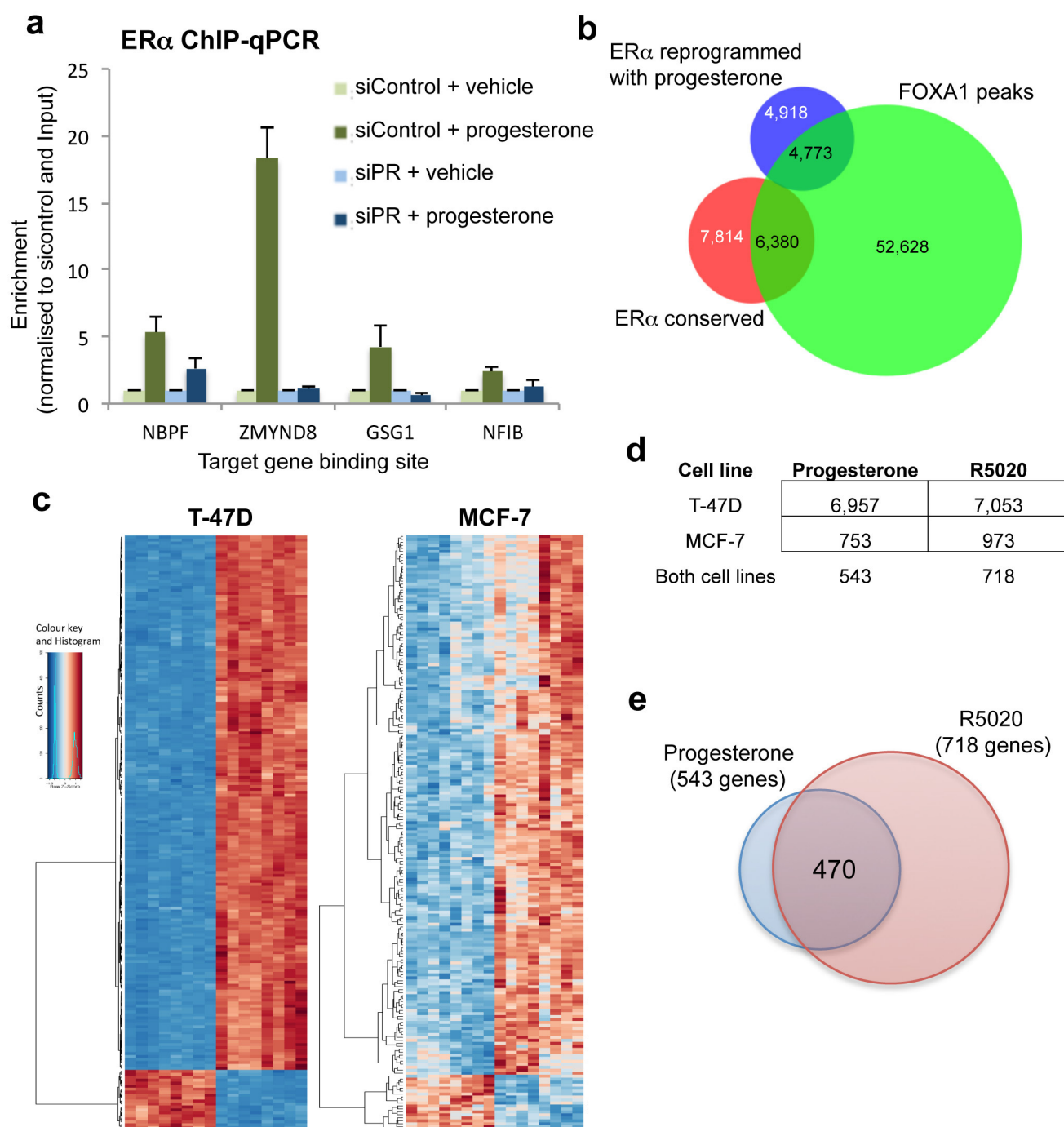
a



b



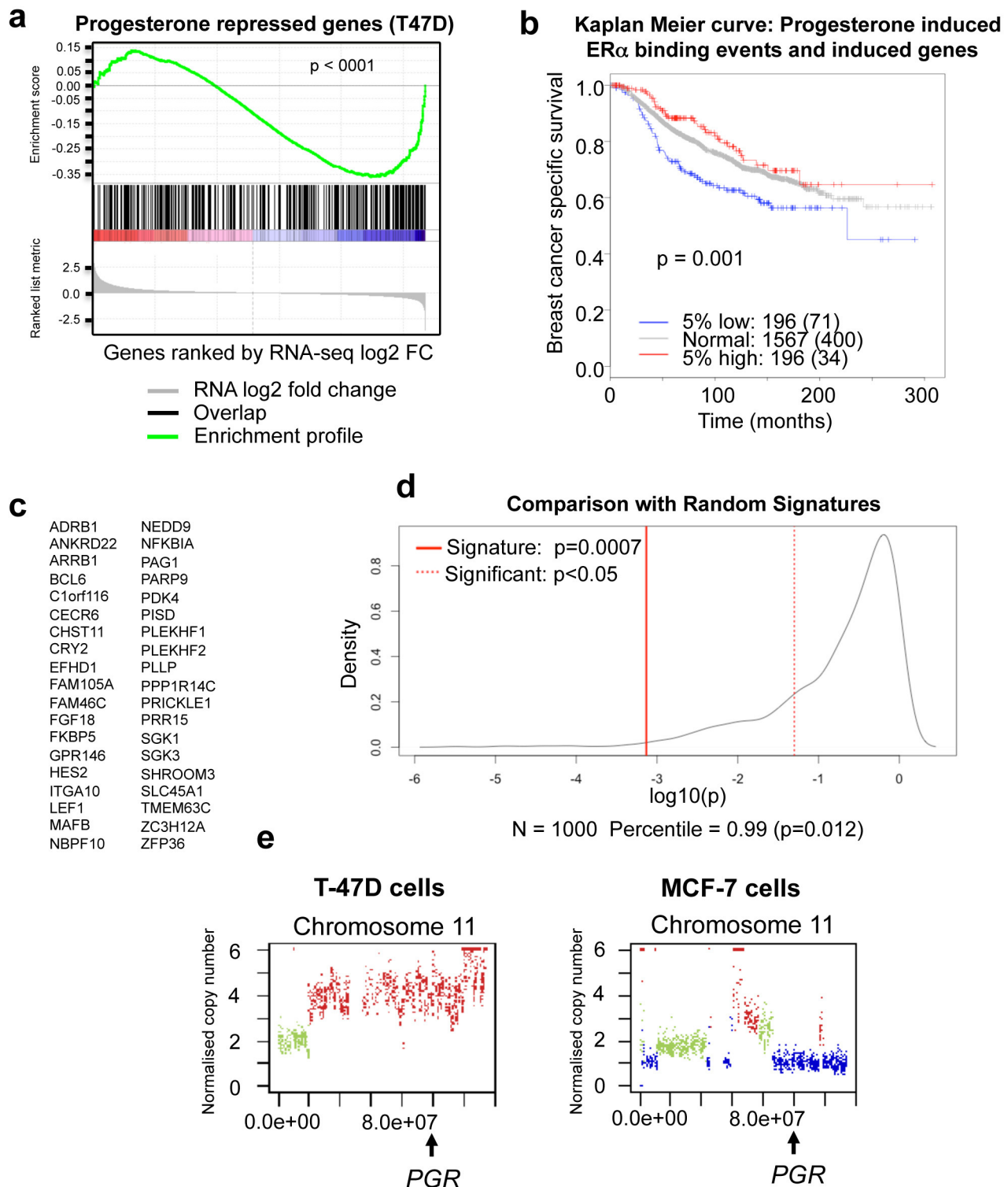
Extended Data Figure 3 | ERα binding in single hormone conditions.
a, b, T47-D (**a**) or MCF-7 (**b**) cells were hormone-deprived and treated with vehicle control, oestrogen alone or progesterone alone. ERα ChIP-seq was conducted and we assessed the binding at the regions previously shown to be reprogrammed by oestrogen plus progesterone. The ERα reprogramming data under both oestrogen and progesterone conditions in the T-47D cells is from Fig. 2b. In the absence of oestrogen, progesterone does not induce ERα binding. In the absence of progesterone, oestrogen does not induce ERα binding to the locations shown to acquire reprogrammed ER binding events under dual hormone conditions.



Extended Data Figure 4 | Validation of binding and gene expression changes.

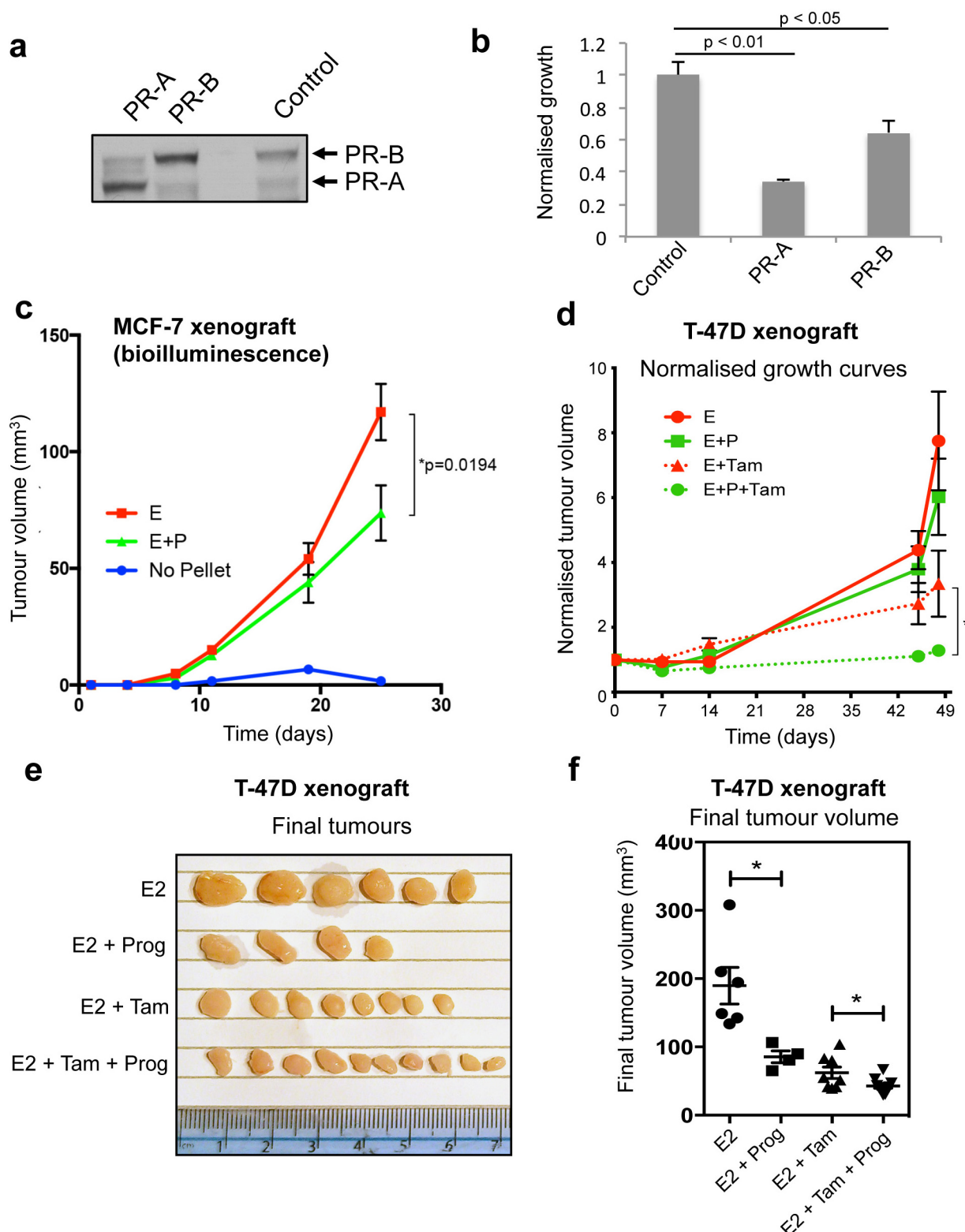
a, Validation of dependence on PR for ER α binding and overlap between ER α binding and FOXA1 binding. T-47D cells were grown in full, oestrogen-rich media and transfected with siControl or siRNA to PR. ER α ChIP was conducted followed by qPCR of several novel ER α binding events only observed under progesterone treatment conditions. In the absence of PR, ER α is not able to associate with the progesterone-induced binding sites. The figure represents one biological replicate of three completed replicates and the error bars represent standard deviation of the technical ChIP-PCR replicates. **b**, Venn diagram showing the ER α binding events that were

conserved in T-47D cells (that is, not altered by progesterone when compared to oestrogen alone) and the ER α binding events that were reprogrammed by progesterone treatment, when overlapped with FOXA1 ChIP-seq data from T-47D cells. The FOXA1 ChIP-seq data from T-47D cells was from ref. 19. **c**, Differential gene changes in MCF-7 and T-47D cells following treatment with progesterone or R5020 for 3 h. Heat map showing gene changes relative to matched controls. Eight replicates were included. **d**, Table showing the differentially regulated genes in the two cell lines and in the two treatment conditions. **e**, Overlap between genes regulated by progesterone (in both cell lines) and gene regulated by the synthetic progestin R5020 (in both cell lines).



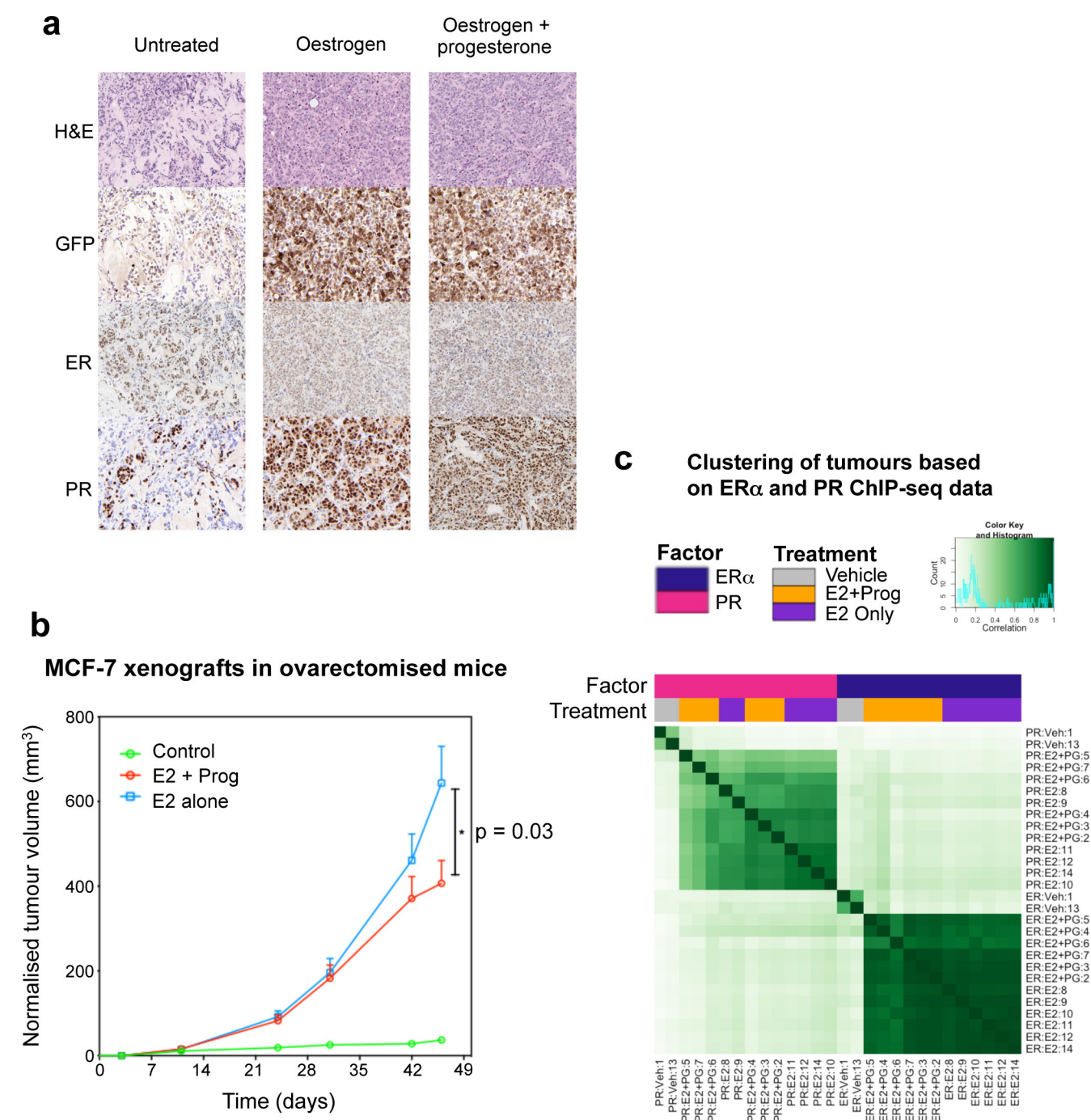
Extended Data Figure 5 | Analysis of gene expression changes and generation of gene signature. **a**, RNA-seq was conducted after progesterone or R5020 treatment for 3 h. GSEA analysis was conducted on progesterone/R5020 repressed genes with lost ER α binding events observed in T47-D cells. The progesterone-decreased ER α binding regions correlate with progesterone downregulated genes. **b**, Kaplan–Meier survival curve in 1,959 breast cancer patients based on a gene signature derived from the progesterone regulated genes and progesterone regulated ER α binding events. **c**, For a gene to be considered it was differentially regulated by progesterone/progestin (as measured by RNA-seq) and the gene had a differentially regulated ER α binding

event within 10 kb of the transcription start site. This resulted in 38 genes. **d**, Performance of progesterone induced gene signature at separating based on survival over 392 patients in top or bottom 10% of expression compared to null distribution of P values computed using 1,000 randomly selected 38-gene signatures. **e**, Copy number alterations on chromosome 11 in T-47D and MCF-7 cells. Green is copy number neutral, blue is copy number loss and red is copy number gain. T-47D cells have an amplification of the chromosome 11 region encompassing the *PGR* gene and MCF-7 cells have a copy number loss of this genomic region.



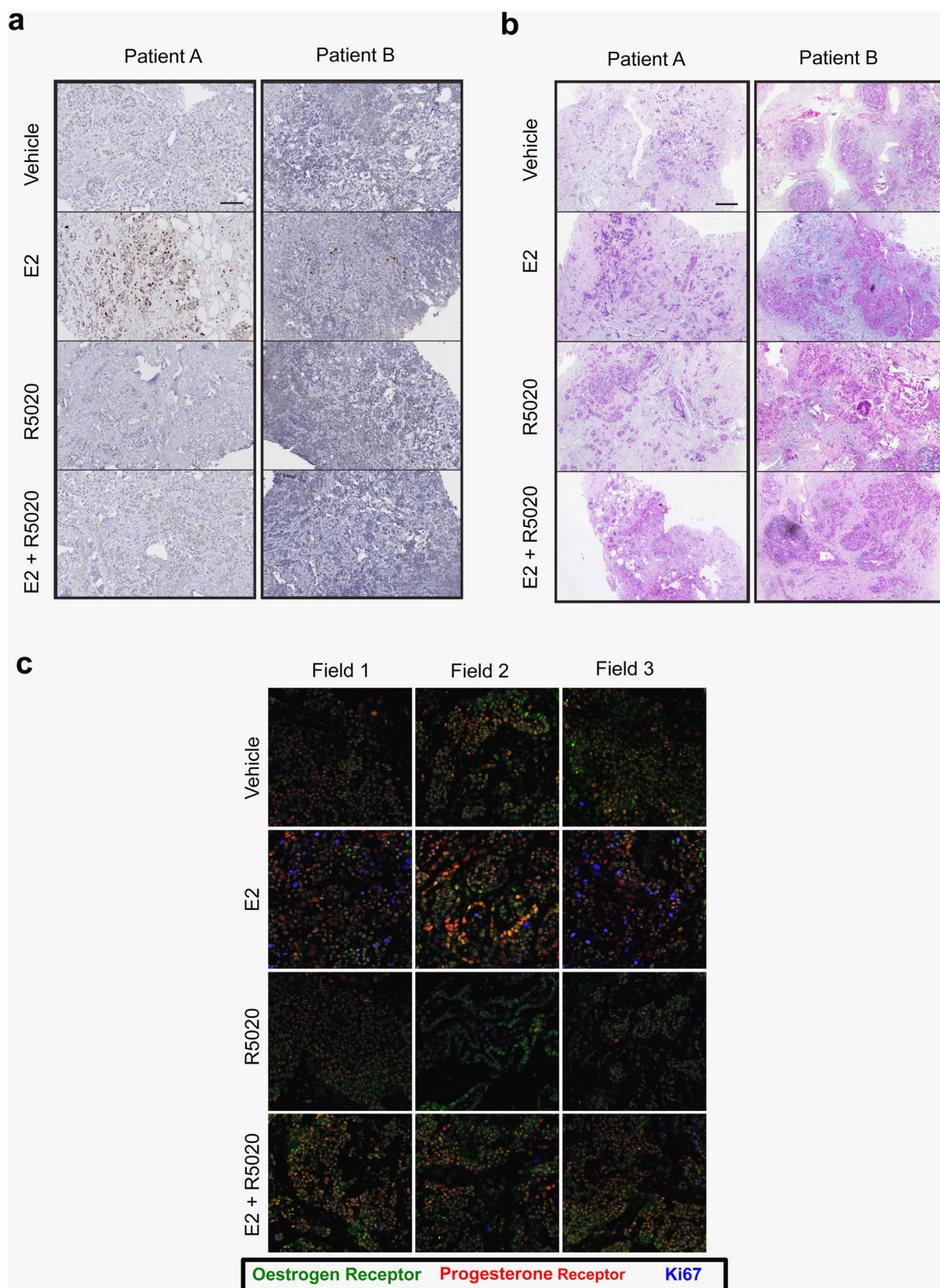
Extended Data Figure 6 | PR inhibits cell line growth and progesterone inhibits T-47D xenograft growth. **a**, MCF-7 cells were transfected with control vector, PR-A or PR-B expressing vectors. Western blotting confirmed the expression of the appropriate PR isoform. **b**, Growth was assessed following oestrogen plus progesterone treatment. The graph represents the average of three independent biological replicates and the error bars represent standard deviation. **c**, Assessment of MCF-7 xenograft tumour growth by physical measurement of tumour volume. Ten tumours for each condition (two in each of five mice per condition) were included. The data were analysed using a *t*-test

and the error bars represent \pm s.e.m. **d**, T-47D xenografts were established in NSG mice. Ten tumours for each condition (two in each of five mice per condition) were included. All were grown in the presence of oestrogen (E2) pellets and subsequently supplemented with vehicle, progesterone, tamoxifen or tamoxifen plus progesterone. Normalized tumour growth is shown. The data were analysed using a *t*-test and the error bars represent \pm s.e.m. **e**, Final T-47D xenograft tumour volumes are shown. **f**, Final T-47D xenograft tumour volumes plotted graphically.



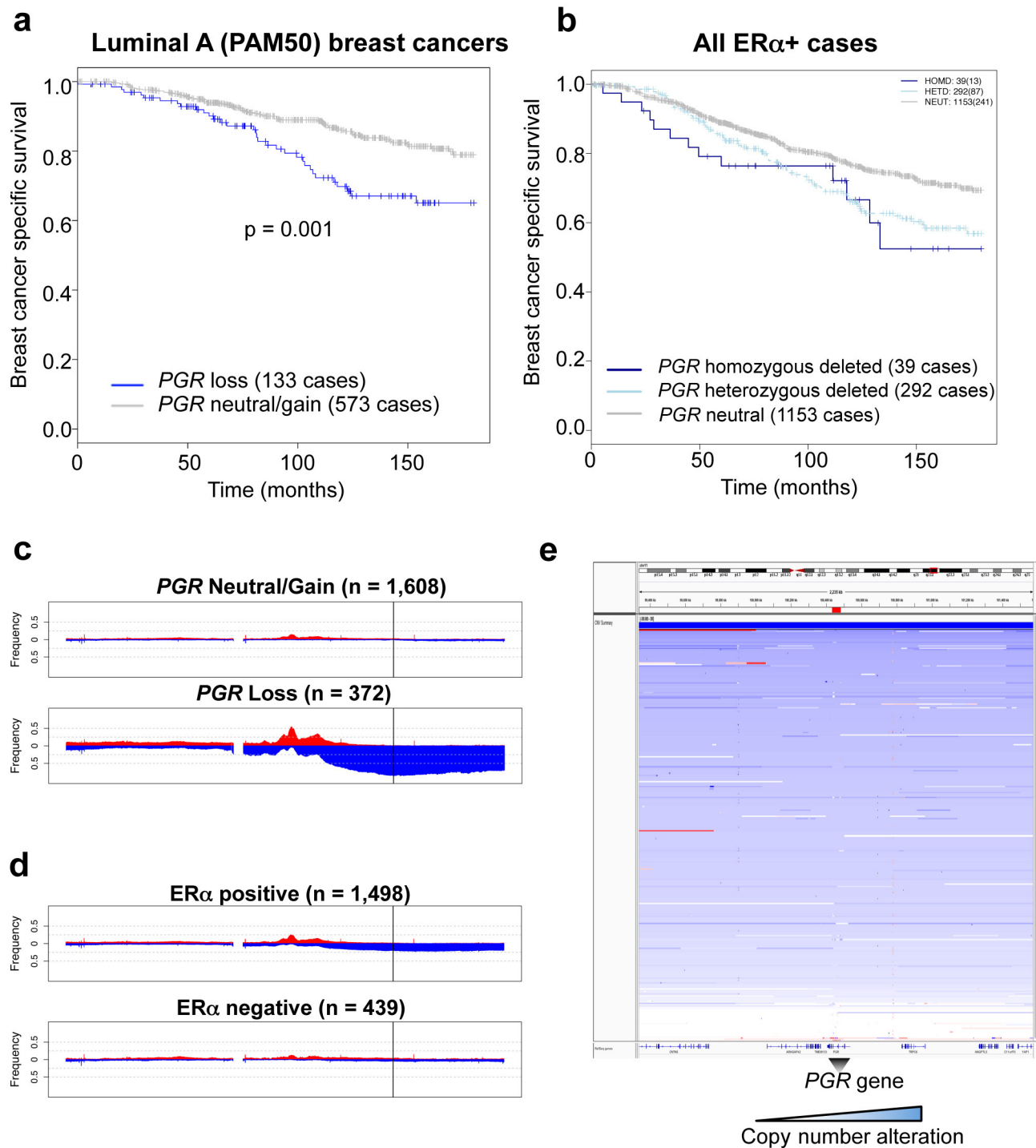
Extended Data Figure 7 | Histological analysis of xenograft tumours and ChIP-seq from xenograft tumours in ovariectomized mice. **a**, Histological analysis of MCF-7 xenograft tumours in untreated, oestrogen or oestrogen plus progesterone conditions. Tumours were taken from 25 day treated conditions. The human xenograft cells expressed GFP, permitting discrimination between human tumour cells and mouse host cells. MCF-7 xenograft experiment in ovariectomized mice. **b**, In order to map ER α binding events by ChIP-seq in MCF-7 xenograft tumours, we repeated the experiment in

ovariectomized mice to eliminate any issues related to the endogenous mouse progesterone. Ten tumours for each condition (two in each of five mice per condition) were included. Growth of xenograft tumours under different hormonal conditions, Control, oestrogen alone (E2) and oestrogen plus progesterone (E2 + Prog). The data were analysed using a *t*-test and the error bars represent \pm s.e.m. **c**, ChIP-seq for ER α and PR were conducted in six matched tumours from each hormonal condition. Also included were two tumours from no hormone conditions. Correlation heat map of all samples.



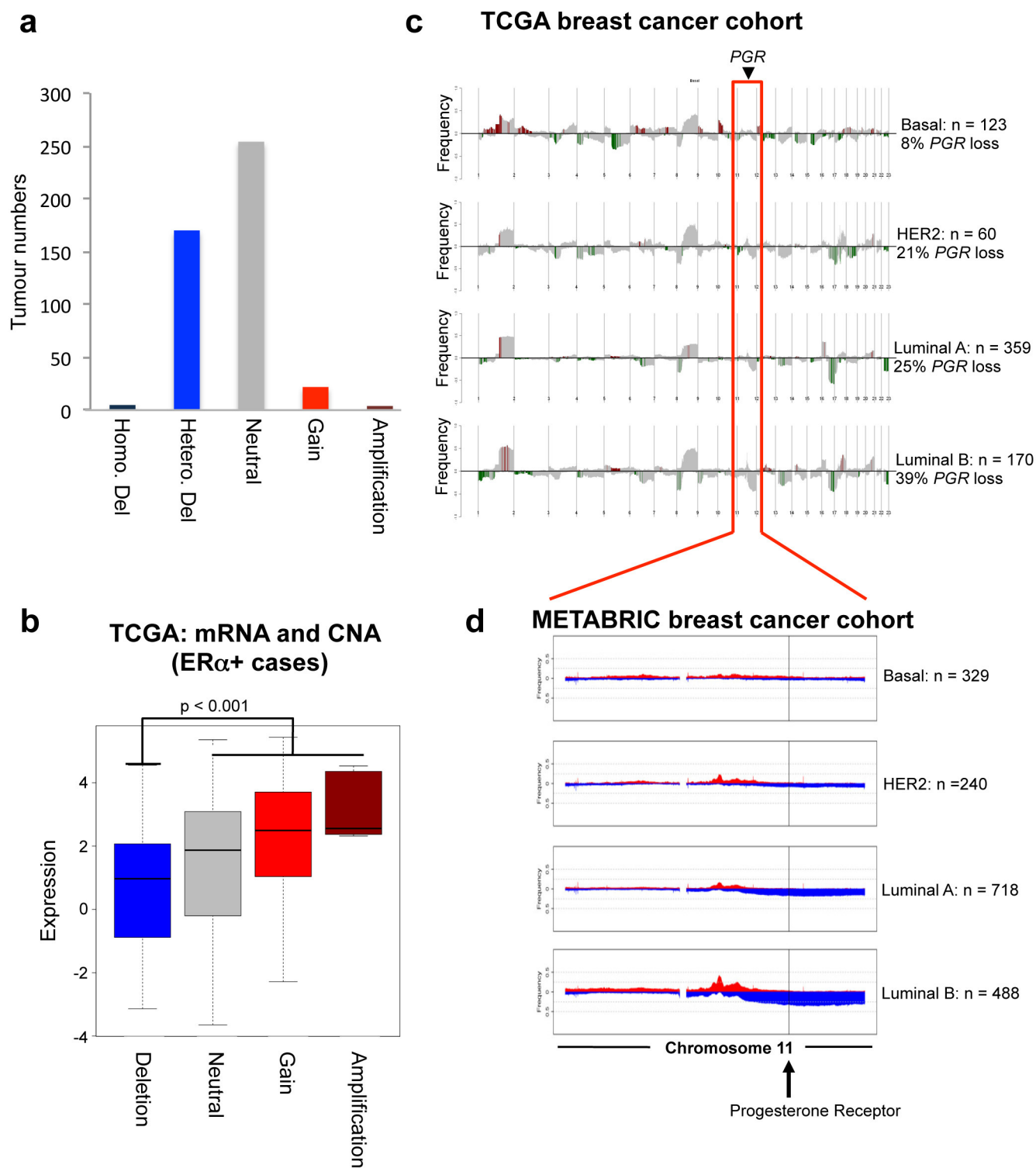
Extended Data Figure 8 | Primary tumours cultivated as *ex vivo* explants shown response to progesterone. Representative images of primary breast cancer explant tissue sections treated with vehicle, oestrogen (E2), the progestin R5020 or oestrogen plus progestin (E2 + R5020). **a**, **b**, These sections were probed with anti-Ki67 (brown) to label proliferating cells (**a**) or haematoxylin and eosin (**b**). Each image is of a single tissue segment from a selection of

3–4 sections per sample treatment. Scale bars, 100 μ m. **c**, Confocal microscopy images (representative fields from each of the triplicate fragments) of a representative primary breast cancer explant tissue treated with vehicle, oestrogen (E2), the progestin R5020 (Progestin) or oestrogen plus progestin (E2 + R5020) and probed with anti-ER α (green), anti-PR (red) and anti-Ki67 to assess proliferating cells (blue).



Extended Data Figure 9 | Analysis of *PGR* copy number loss in the METABRIC cohort. **a**, Kaplan–Meier analysis of breast cancer specific survival within the METABRIC cohort. Only within luminal A tumours (based on PAM50 gene expression signature), tumours were stratified based on copy number loss of *PGR* or not. In total 19% of luminal A tumours contain a copy number loss of the *PGR* genomic locus and these patients have a poorer clinical outcome. **b**, All ER α + cases were stratified based on *PGR* copy number status, showing tumours with heterozygous and homozygous deletions

separately. **c**, Chromosome 11 in tumours with neutral or gained *PGR* versus those with copy number loss of the *PGR* gene (defined by line). **d**, Chromosome 11 copy number status between ER α positive and negative tumours. **e**, Visual representation of all ER α + tumours with a copy number alteration at the *PGR* genomic locus, showing the copy number changes relative to the *PGR* gene (highlighted below) and the surrounding ~2.2 Mb of genomic sequence.



Extended Data Figure 10 | Validation of genomic copy number loss in the *PGR* gene in an independent data set. **a**, TCGA ER α ⁺ breast cancers were assessed for copy number changes in *PGR*. The number of tumours in each category, based on copy number changes. Only ER α ⁺ breast cancers were included. **b**, Correlation between PR mRNA levels and copy number status in all luminal breast cancers within the TCGA cohort. The heterozygous and

homozygous deletions are combined. **c**, Frequency of copy number alterations across entire genome in TCGA breast cancer cohort, stratified based on subtype using PAM50 signature. Chromosome 11, which encompasses *PGR* gene is highlighted and the frequency of copy number loss of the *PGR* genomic region is provided. **d**, Copy number changes on chromosome 11 within the METABRIC cohort, based on subtype stratification (PAM50 signature).

Rapidly rotating second-generation progenitors for the ‘blue hook’ stars of ω Centauri

Marco Tailo^{1,2}, Francesca D’Antona¹, Enrico Vesperini³, Marcella Di Criscienzo¹, Paolo Ventura¹, Antonino P. Milone⁴, Andrea Bellini⁵, Aaron Dotter⁴, Thibaut Decressin¹, Annibale D’Ercole⁶, Vittoria Caloi⁷ & Roberto Capuzzo-Dolcetta²

Horizontal branch stars belong to an advanced stage in the evolution of the oldest stellar galactic population, occurring either as field halo stars or grouped in globular clusters. The discovery of multiple populations in clusters^{1,2} that were previously believed to have single populations gave rise to the currently accepted theory that the hottest horizontal branch members (the ‘blue hook’ stars, which had late helium-core flash ignition³, followed by deep mixing^{4,5}) are the progeny of a helium-rich ‘second generation’ of stars^{6,7}. It is not known why such a supposedly rare event^{8,9} (a late flash followed by mixing) is so common that the blue hook of ω Centauri contains approximately 30 per cent of the horizontal branch stars in the cluster¹⁰, or why the blue hook luminosity range in this massive cluster cannot be reproduced by models. Here we report that the presence of helium core masses up to about 0.04 solar masses larger than the core mass resulting from evolution is required to solve the luminosity range problem. We model this by taking into account the dispersion in rotation rates achieved by the progenitors, whose pre-main-sequence accretion disk suffered an early disruption in the dense environment of the cluster’s central regions, where second-generation stars form¹¹. Rotation may also account for frequent late-flash-mixing events in massive globular clusters.

In the colour-magnitude diagrams of globular clusters, the horizontal branch is the locus of core-helium-burning structures that are the progeny of red giant stars. For each such structure, the helium flash ignition has occurred either at the tip of the red giant branch, or anywhere along the evolutionary path that moves the star into the white dwarf stage⁵. In general no mixing occurs, and the result is a ‘standard’ horizontal branch structure: the smaller its hydrogen-rich envelope is, the hotter is the location of effective temperature T_{eff} in the model in the Hertzsprung–Russell diagram (that is, along the horizontal branch), up to about 32,000 K. A very late helium ignition, along the white-dwarf cooling pathway, can cause the helium core to mix with the small hydrogen-rich envelope^{5,12,13} (late-flash-mixing), resulting in a slightly smaller helium-burning core plus a helium-rich, or a helium-dominated, envelope. Such ‘blue hook’ structures attain a lower luminosity and a higher T_{eff} than do stars in the ‘extreme’ hottest horizontal branch standard locus, owing to the smaller opacity of the helium-rich atmosphere; they have been found in the ω Cen cluster¹⁴ and in a few other massive clusters⁹.

The blue hook in ω Cen is particularly striking because it contains about 30% of the horizontal branch stars¹⁰ and is extremely well defined in the Hubble Space Telescope observations. The optical colour-magnitude diagram¹⁰ displays a blue hook that extends approximately 1.0 magnitudes in the F625W band, with a redder side that is much less populated (Extended Data Fig. 1a). A smaller sample of blue and ultraviolet data¹⁵ (Fig. 1a) displays a strong peak on the blue

side of the colour distribution, where the helium-rich late-flash-mixed stars are located^{16,17}.

The models we adopt (see Methods) assume the helium core masses implied by the evolutionary process, and different initial envelope hydrogen abundances ($X_{\text{env-in}}$), together with an efficiency of helium settling calibrated to reproduce as well as possible the values of He/H versus T_{eff} data from the literature¹⁷ (Fig. 2). Tracks (that is, evolutionary sequences) with very low $X_{\text{env-in}}$ values span only about 0.4 mag in F225W, but an extra 0.35 mag or so are obtained in models starting with larger $X_{\text{env-in}}$ values, as a result of the change in atmospheric composition from helium-dominated to hydrogen-rich (Extended Data Fig. 2 and Extended Data Table 1). Considering the whole range of $X_{\text{env-in}}$ values explored, from 0.007 to 0.41, we can obtain a global extension of 0.9 mag in the F225W band (6th column of Extended Data Table 1), close to the range observed, but the tracks with $X_{\text{env-in}} > 0.06$ do not match the colour-magnitude diagram well (Extended Data Fig. 4), nor the He/H versus T_{eff} data (Fig. 2), and tend to merge with the extreme horizontal branch. For $X_{\text{env-in}} \leq 0.06$, the tracks cover a maximum of 0.69 mag and the total extension is 0.79 mag.

Making the currently accepted assumption that the blue hook contains the progeny of the very-helium-rich second-generation stars, the existence of the blue hook stars can be attributed to late-flash-mixing events in models having an initial helium mass fraction of $Y = 0.37$. Simulations including only objects having up to $X_{\text{env-in}} = 0.06$ show the discrepancy with the observed luminosity range (Fig. 1b).

Modifications in model inputs have not much effect on the covered magnitude range: (1) standard flash-mixed models occur in a small core mass range ($\delta M_c < 0.008 M_\odot$)^{8,9}, corresponding to a negligible magnitude difference in track extension⁹; (2) the magnitude range cannot be extended by increasing the adopted core overshooting parameters; and (3) we also have good reason to reject the possibility that the blue hook contains both the progeny of the standard and of the helium-rich populations (see discussion in Methods and Extended Data Fig. 3).

The high-luminosity portion of the blue hook, at the correct colour location, can be explained by tracks having larger helium core masses M_c , as we show in Fig. 1a and Extended Data Fig. 1a for the track having $\delta M_c = +0.04 M_\odot$ and $Y = 0.37$, where M_\odot is the solar mass. The helium flash ignites in more massive cores if very rapid core rotation delays the attainment of the flash temperatures. Of the (few) models available, we select a first-order approximation¹⁸ of models starting from solid-body rotation on the main sequence, and preserving angular momentum in shells, to estimate the increase in the helium-flash core mass, δM_c , as a function of the initial angular velocity ω . These models provide upper limits to δM_c , because they do not allow for angular momentum transport and losses during main-sequence and post-main-sequence evolution. In this approximation,

¹INAF-Osservatorio Astronomico di Roma, I-00040 Monte Porzio (Roma), Italy. ²Dipartimento di Fisica, Università degli Studi di Roma ‘La Sapienza’, Piazzale Aldo Moro 5, I-00185 Roma, Italy. ³Department of Astronomy, Indiana University, Bloomington, Indiana, 47405-7105, USA. ⁴Research School of Astronomy and Astrophysics, Australian National University, Canberra, Australian Capital Territory 2611, Australia. ⁵Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ⁶INAF-Osservatorio Astronomico di Bologna, via Ranzani 1, I-40127 Bologna, Italy. ⁷INAF, IAPS, Roma, via Fosso del Cavaliere 100, I-00133 Roma, Italy.

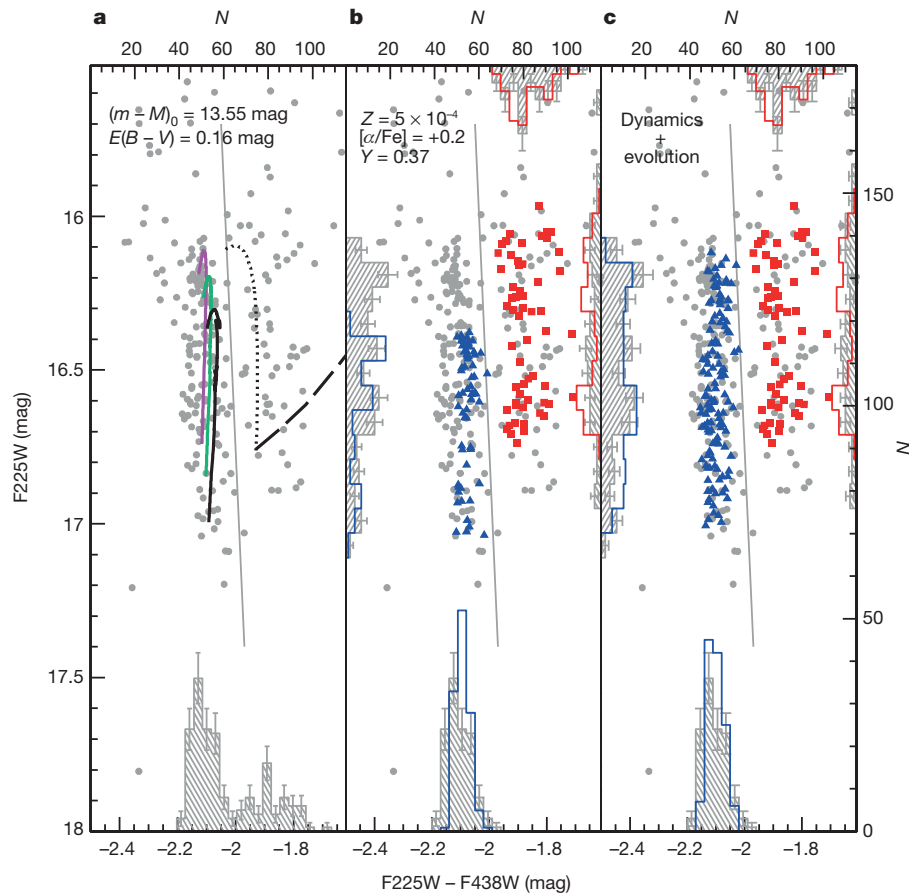


Figure 1 | Simulations for the ultraviolet data. Grey dots are the observed stars in ω Cen. Tracks plotted on the blue hook in **a** are: for $M = 0.466M_{\odot}$ and $Y = 0.37$ (black); for $0.489M_{\odot}$, $Y = 0.25$, standard M_c (green); and for $M = 0.466M_{\odot}$ and $Y = 0.37$, M_c increased by $0.04M_{\odot}$ (magenta). Models have $X_{\text{env-in}} = 0.06$. Also shown are the zero-age horizontal branch for $Y = 0.37$ (dashed black) and the $0.471M_{\odot}$ track (dotted black). Labels indicate bolometric distance modulus $(m - M)_0$ and reddening $E(B - V)$ adopted

for the comparison. **b**, Simulation with a late-flash mixture of $X_{\text{env-in}}$ stars for $Y = 0.37$ (see text for the details of the chemistry, labels at the top). Comparisons with data are shown in the histograms on the left (blue hook, blue triangles) and right (extreme horizontal branch, red squares) side. On the histograms, 1σ Poisson errors are shown with respect to the counts N . **c**, Simulation resulting from coupling dynamics and evolution (same mixture of $X_{\text{env-in}}$). Colour code and comparisons as in **b**.

about half of the break-up main-sequence rotation rate (at which the centrifugal force at the equator becomes equal to the gravitational force, which is about $7 \times 10^{-4} \text{ s}^{-1}$) provides a huge $\delta M_c = 0.06M_{\odot}$, leaving room for different $M_c(\omega)$ laws when available.

We propose that high rotation rates may be a consequence of the star-formation history and early dynamics in very massive clusters. Specifically, in the model based on the formation of second-generation stars from the ejecta of asymptotic giant branch stars, a cooling flow collects such ejecta in the central regions of the first-generation cluster and produces a centrally concentrated second-generation subcluster^{11,19}. Observations showing that the helium-rich blue main-sequence stars—the progenitors of the blue hook stars—are spatially more concentrated than the rest of the main-sequence stars²⁰ and retain some ‘memory’ of their initial spatial segregation, provide evidence that indeed such progenitors formed segregated in the innermost regions of the cluster, as predicted¹¹ and assumed in this study.

The contracting pre-main-sequence low-mass objects in the galactic disk (that is, stars like the variable star T Tauri) rotate with periods in the range of 2 days to 12 days. The rotation rates of classical T Tauri stars do not increase with stellar age, because of magnetic disk-locking between the star and the disk²¹, which keeps the stellar rotation constant until the disk is lost. Stars in young clusters (such as α Persei) show a wide distribution of rotation velocities, owing to differences in the time at which different objects break their magnetic coupling with the wind²². Rapidly rotating stars result from an early breaking of the

coupling. Afterwards, during the main sequence, stellar winds slow down the rotation of the convective stellar envelopes, and old stars all appear to rotate slowly, but the inner core rotation is still fast despite the slow angular momentum transfer from the core to the envelope. We note that this model implies the presence of a crowded second generation.

Figure 3a shows the fast decrease of the momentum of inertia of first- and second-generation stars in globular clusters (the latter ones have no initial deuterium, because the gas from which they formed was nuclearily processed at high temperature¹¹). Figure 3b describes the stellar angular velocity which can be obtained at the main sequence (age exceeding 2×10^7 yr), as a function of the time at which the disk-star coupling is destroyed and evolution proceeds at constant angular momentum. The moment of inertia evolution of the case of a mass of $M = 0.7M_{\odot}$, $Y = 0.35$ is assumed, for rotation periods at detachment taken in the range 2–12 days. We see that there is ample model space to reach high angular velocities, provided that the second-generation stars (the progenitors of the blue hook stars) lose their disks at young ages, from 10^5 yr (for the longer initial periods) to about 3×10^6 yr (for the short periods).

We model the dynamical encounters that destroy the disk, assuming that three encounters are necessary (see details in Methods), and use the δM_c resulting from the timings of encounters as direct inputs for the simulation. The resulting distribution of the blue hook is shown in Fig. 1c and Extended Data Fig. 1b. We remark that fast rotation will

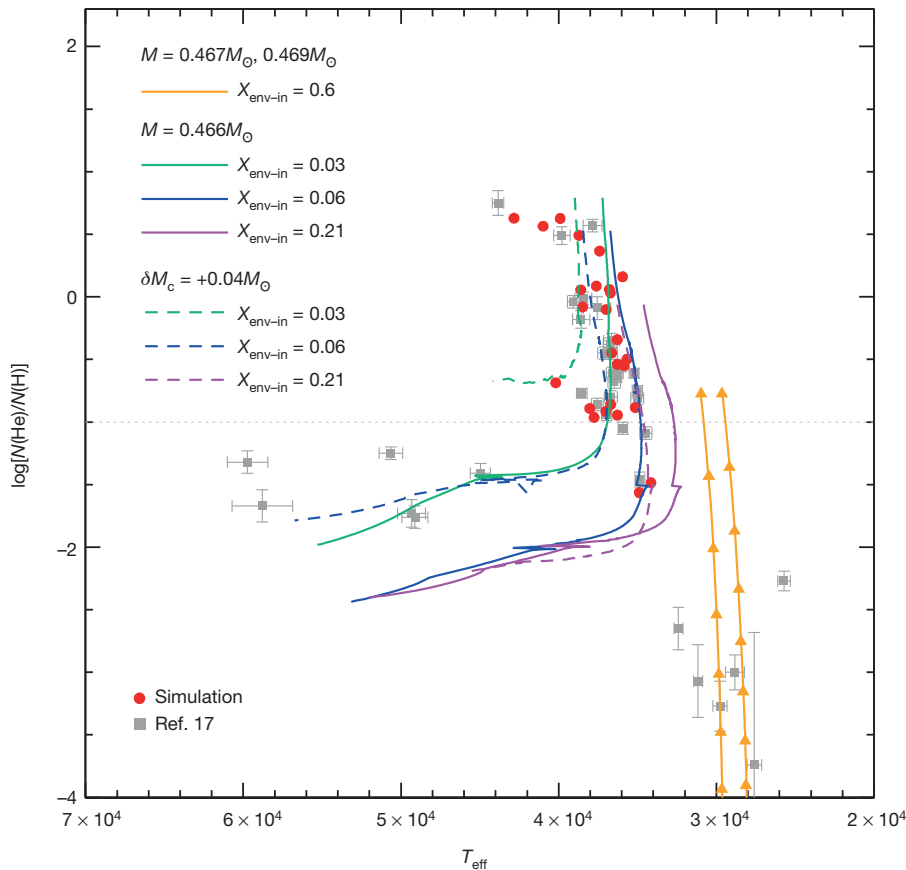


Figure 2 | Calibration of hydrogen-helium diffusion. Grey squares show the He/H versus T_{eff} data¹⁷ with error bars (for definition of the error bars, see ref. 17). The solar ratio is shown as a horizontal grey dotted line. The evolution of three models is shown for different values of $X_{\text{env-in}}$. Diffusion is computed assuming that the envelope mass below which we assume diffusion is operating, M_{turb} , is $10^{-6} M_{\odot}$. The orange tracks show the evolution of the smallest non-mixed tracks. Triangles mark each 10^7 -yr interval, for $M_{\text{turb}} = 10^{-7}$. The red dots represent 26 stars randomly extracted from the simulation in Fig. 1c. Along the magenta curve, no star with $X_{\text{env-in}} = 0.21$ is extracted because such a star would be cooler than the observed sample.

favour deep mixing in the giant envelopes during the last phases of evolution, enhancing the chemical anomalies of the most extreme second generation^{23,24} and probably also favouring mixing at the onset of the flash: the late-flash-mixing event, although very different from what is described in one-dimensional models, will be more probable, thanks to the reduced entropy barrier between the core and the envelope, justifying the existence of the blue hook itself. In this scheme, the ‘non-mixed’ extreme horizontal branch stars should have been slowly rotating, and in fact standard models match this group well. In the simulation we technically model them by assuming that the cool side of the hot horizontal branch is populated by the progeny of scarcely rotating (rotation rate $\omega < 10^{-6} \text{ s}^{-1}$) stars (red squares in Fig. 1 and Extended Data Figs 1, 3 and 4). M_c for this side of the simulation is the mass of the non-rotating core, and in fact standard tracks match the extreme horizontal branch well.

The presence of rapidly rotating stars among the second-generation population should not be confined to modelling the blue hook in ω Cen. Two conditions are at the basis of the success of the dynamical model used here: first, we must deal with second-generation stars, which form in a cooling flow at stellar densities much higher than that of the first-generation stars; and second, the population must be abundant, so it requires the presence of very massive clusters to start with. Of the most massive clusters, M54 (NGC 6715) shows a blue hook very similar in shape to that of ω Cen²⁵, and NGC 2419 also has an extended blue hook²⁶. NGC 6388 and NGC 6441 have peculiarly extended horizontal branches, unlike other metal-rich clusters. It is certainly possible that their thick red horizontal branch (the ‘red clump’), modelled by assuming a large helium spread in the second generation²⁷, is also caused, in part, by the presence of larger core masses due to fast rotation. The reproduction of the horizontal branch morphology would then require, in the second generation, a smaller helium content increase than predicted by standard models²⁷, and this

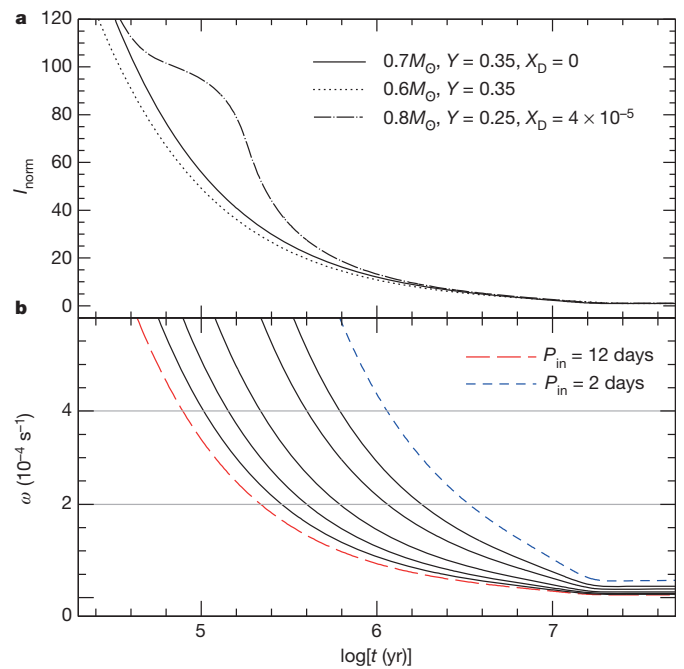


Figure 3 | Rotational evolution from disk detachment to the main sequence. **a**, Time evolution of moment of inertia (I_{norm} , in units of moment at $t = 2 \times 10^7$ yr), for different masses, values of Y and deuterium abundance X_D . **b**, For $M = 0.7 M_{\odot}$ and $Y = 0.35$, we show the rotation rate ω at age $t = 2 \times 10^7$ yr, as a function of the age at which the star detaches from the proto-stellar disk, and conservation of angular momentum begins. The lines represent different assumptions for the stellar rotation period when the disk is lost, from left to right: $P_{\text{in}} = 12$ days, 10 days, 8 days, 6 days, 4 days, 3 days and 2 days.

would be more consistent with the helium spread derived by the main-sequence colour thickness in NGC 6441²⁸. We note that a large rotation spread in this case does not produce a prominent blue hook, but an anomalous red clump. NGC 2808, just a bit less massive, could be a borderline case, in which the presence of fast rotators is uncertain.

Smaller but important rotation ranges may be present in other less-massive clusters as well, and may have been due to the same process of early disk destruction in second-generation stars: in many clusters the mass loss necessary to explain the location of second-generation stars has to be slightly larger than in the first generation, possibly because of their rotation rate^{29,30}.

The connection between the magnitude extension of the blue hook stars in ω Cen and the range of rotation rates of their progenitors might be another manifestation of the interplay between the formation of multiple populations in clusters, their dynamical behaviour and the properties of the component stars. At the same time, the model developed in this work for the blue hook supports the model for the dynamical and chemical formation of multiple populations¹¹ based on the contribution of asymptotic giant branch stars and adopted in this study.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 October 2014; accepted 28 April 2015.

Published online 22 June 2015.

- Gratton, R. G., Carretta, E. & Bragaglia, A. Multiple populations in globular clusters. Lessons learned from the Milky Way globular clusters. *Astron. Astrophys. Rev.* **20**, 50 (2012).
- Piotto, G. *et al.* Hubble Space Telescope reveals multiple sub-giant branch in eight globular clusters. *Astrophys. J.* **760**, 39 (2012).
- D'Cruz, N. L., Dorman, B., Rood, R. T. & O'Connell, R. W. The origin of extreme horizontal branch stars. *Astrophys. J.* **466**, 359 (1996).
- Moehler, S., Sweigart, A. V., Landsman, W. B. & Heber, U. Hot HB stars in globular clusters—physical parameters and consequences for theory. V. Radiative levitation versus helium mixing. *Astron. Astrophys.* **360**, 120–132 (2000).
- Brown, T. M., Sweigart, A. V., Lanz, T., Landsman, W. B. & Hubeny, I. Flash mixing on the white dwarf cooling curve: understanding hot horizontal branch anomalies in NGC 2808. *Astrophys. J.* **562**, 368–393 (2001).
- D'Antona, F., Caloi, V., Montalbán, J., Ventura, P. & Gratton, R. Helium variation due to self-pollution among Globular Cluster stars. Consequences on the horizontal branch morphology. *Astron. Astrophys.* **395**, 69–75 (2002).
- Brown, T. M. *et al.* Flash mixing on the white dwarf cooling curve: spectroscopic confirmation in NGC 2808. *Astrophys. J.* **748**, 85 (2012).
- Miller Bertolami, M. M., Althaus, L. G., Unglaub, K. & Weiss, A. Modeling He-rich subdwarfs through the hot-flasher scenario. *Astron. Astrophys.* **491**, 253–265 (2008).
- Brown, T. M. *et al.* The blue hook populations of massive globular clusters. *Astrophys. J.* **718**, 1332–1344 (2010).
- Anderson, J. & van der Marel, R. P. New limits on an intermediate-mass black hole in Omega Centauri. I. Hubble Space Telescope photometry and proper motions. *Astrophys. J.* **710**, 1032–1062 (2010).
- D'Ercole, A., Vesperini, E., D'Antona, F., McMillan, S. L. W. & Recchi, S. Formation and dynamical evolution of multiple stellar generations in globular clusters. *Mon. Not. R. Astron. Soc.* **391**, 825–843 (2008).
- Sweigart, A. V. Helium mixing in globular cluster stars. In *Third Conf. on Faint Blue Stars* (eds Philip, A. G. D., Liebert, J., Saffer, R. & Hayes, D. S.) 3–27 (L. Davis Press, 1997).
- Cassisi, S., Schlattl, H., Salaris, M. & Weiss, A. First full evolutionary computation of the helium flash-induced mixing in population II stars. *Astrophys. J.* **582**, L43–L46 (2003).
- Whitney, J. H. *et al.* Far-ultraviolet photometry of the globular cluster omega Cen. *Astron. J.* **108**, 1350–1363 (1994).
- Bellini, A. *et al.* A double white-dwarf cooling sequence in ω Centauri. *Astrophys. J.* **769**, L32 (2013).
- Moehler, S. *et al.* The hottest horizontal-branch stars in ω Centauri. Late hot flasher vs. helium enrichment. *Astron. Astrophys.* **475**, L5–L8 (2007).
- Latour, M. *et al.* A helium-carbon correlation on the extreme horizontal branch in ω Centauri. *Astrophys. J.* **795**, 106 (2014).
- Mengel, J. G. & Gross, P. G. Possible effects of internal rotation in low-mass stars. *Astrophys. Space Sci.* **41**, 407–415 (1976).
- Bekki, K. Rotation and multiple stellar population in globular clusters. *Astrophys. J.* **724**, L99–L103 (2010).
- Bellini, A. *et al.* Radial distribution of the multiple stellar populations in ω Centauri. *Astron. Astrophys.* **507**, 1393–1408 (2009).
- Armitage, P. J. & Clarke, C. J. Magnetic braking of T Tauri stars. *Mon. Not. R. Astron. Soc.* **280**, 458–468 (1996).
- Bouvier, J., Forestini, M. & Allain, S. The angular momentum evolution of low-mass stars. *Astron. Astrophys.* **326**, 1023–1043 (1997).
- Weiss, A., Denissenkov, P. A. & Charbonnel, C. Evolution and surface abundances of red giants experiencing deep mixing. *Astron. Astrophys.* **356**, 181–190 (2000).
- D'Antona, F. & Ventura, P. A model for globular cluster extreme anomalies. *Mon. Not. R. Astron. Soc.* **379**, 1431–1441 (2007).
- Piotto, G. *et al.* The Hubble Space Telescope UV Legacy Survey of Galactic Globular Clusters. I. Overview of the project and detection of multiple stellar populations. *Astron. J.* **149**, 9 (2015).
- Di Criscienzo, M. *et al.* An HST/WFC3 view of stellar populations on the horizontal branch of NGC 2419. *Mon. Not. R. Astron. Soc.* **446**, 1469–1477 (2015).
- Caloi, V. & D'Antona, F. NGC 6441: another indication of very high helium content in globular cluster stars. *Astron. Astrophys.* **463**, 949–955 (2007).
- Bellini, A. *et al.* The intriguing stellar populations in the globular clusters NGC 6388 and NGC 6441. *Astrophys. J.* **765**, 32 (2013).
- Salaris, M., Cassisi, S. & Pietrinferni, A. The horizontal branch of NGC 1851: constraints on the cluster subpopulations. *Astrophys. J.* **678**, L25–L28 (2008).
- D'Antona, F. *et al.* The puzzle of metallicity and multiple stellar populations in the globular clusters in Fornax. *Mon. Not. R. Astron. Soc.* **434**, 1138–1150 (2013).

Acknowledgements A.P.M. acknowledges support by the Australian Research Council through Discovery Early Career Researcher Award DE150101816. E.V. acknowledges support from grant NASANNX13AF45G. P.V. and F.D'A. acknowledge support from PRIN INAF 2011 “Multiple populations in globular clusters: their role in the Galaxy assembly” (principal investigator E. Carretta), and P.V. acknowledges support from PRIN MIUR 2010–2011, project “The Chemical and Dynamical Evolution of the Milky Way and Local Group Galaxies” (principal investigator F. Matteucci). T.D. acknowledges support from the UE Program (FP7/2007–2013) under grant agreement number 267251 of Astronomy Fellowships in Italy (ASTROFit). M.D.C. acknowledges support from INAF-OAR. A.B. acknowledges support from STScI grant AR-12656.

Author Contributions M.T., F.D'A., E.V. and M.D.C. jointly designed and coordinated this study. F.D'A. proposed and designed the rotational evolution model. E.V. designed and computed the dynamical simulation. M.T. and P.V. computed the new evolutionary models. A.D. computed synthetic colours with an internally consistent treatment of extinction across all bandpasses. M.T. and M.D.C. performed the simulations and the analysis. A.B. performed the data reduction and calibration for the WFC3/UVIS exposures. A.M. dealt with the optical data and the comparison with spectroscopic data. T.D. dealt with the problems connected to the modelling of stellar rotation. V.C. contributed to the discussion and to the writing of the text. A.D'E. and R.C.-D. provided insight on the dynamical aspects. All authors read, commented on and approved submission of this article.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.D'A. (franca.dantona@gmail.com).

METHODS

The data sets. We analyse optical¹⁰ and ultraviolet¹⁵ Hubble Space Telescope data for the hot horizontal branch stars of ω Cen. Photometric errors are about 0.015 mag in filters F435W and F625W of the ACS/WFC (Wide Field Channel of the Advanced Camera for Surveys) and 0.010 mag in the F225W and F438W filters of the WFC3. In the comparisons with optical data, we adopt distance moduli and reddening that fit the luminous part of the horizontal branch in the same data (M.T., PhD thesis in preparation). The values correspond to $E(B - V) = 0.156$ and $(m - M)_0 = 13.63$, when using the extinction ratios $A_{F435W}/A_V = 1.362$ and $A_{F625W}/A_V = 0.868$, explicitly computed for this work, and appropriate for stars of $T_{\text{eff}} > 20,000$ K, using standard extinction curves^{31,32}, and $A_V = 3.1 \times E(B - V)$. The F225W and F438W data¹⁵ have been recalibrated. Distance modulus and reddening are chosen by adjustment of the simulations results, and are labelled in Fig. 1. Using $A_{F225W}/A_V = 2.670$ and $A_{F438W}/A_V = 1.356$, the bolometric distance modulus turns out to be about 0.08 mag smaller than the modulus derived from optical data.

The prominent peak shown by number counts on the blue side (histogram at the bottom of Fig. 1a), followed by a sharp decline, and by a second broader peak at redder colours, corresponds to a change in the spectroscopic abundances: the cool side of the distribution shows very helium-poor spectra¹⁶, a feature ascribed to the effect of helium-diffusion in hydrogen-rich envelopes (the hottest standard horizontal branch stars). An almost abrupt transition to hydrogen–helium intermediate composition occurs^{16,17} at $T_{\text{eff}} \approx 35,000$ K, where the sharp peak occurs. The contemporary presence of helium and carbon^{7,17,33} means that the stars at $T_{\text{eff}} \geq 35,000$ K are the result of late-flash–mixing⁵, so they must be blue hook stars. Standard one-dimensional simulations of late-flash–mixing end up with a very small hydrogen abundance left in the envelope ($X \approx 4 \times 10^{-4}$; see ref. 13). A ‘shallow’ mixing, leaving some percentage of hydrogen abundance, occurs only for a metal mass fraction that is ten times larger ($Z \geq 0.01$; ref. 8) than the relevant metallicity in our sample. Nevertheless, the comparison with observed number ratio values $N(\text{He})/N(\text{H})$ shows that many blue hook stars must be the result of shallow mixing.

We hand-draw a possible dividing line between blue hook and extreme, non-mixed, horizontal branch stars. With this subdivision, in the optical sample there are about 320 blue hook stars and about 150 cooler stars. In the ultraviolet sample we get 130 blue hook stars plus 60 cooler stars.

Standard models. Horizontal branch models are computed with the ATON code³⁴. Some useful results are given in Extended Data Tables 1 and 2. Most models start from the zero-age horizontal branch, where the helium core mass is fixed by previous evolution up to the helium flash, and the rest of the mass is in the hydrogen-rich envelope. We also use as guidelines the results of some full late-flash evolutions, computed following the standard methodology^{3,35} of evolving the mass corresponding to an age of 12 billion years (Gyr) along the red giant branch, for increasing wind mass loss rates. Our results substantially confirm the previous findings^{5,8,13}. The mass range of late-flash stars can be estimated to be $\sim 0.02\text{--}0.03M_{\odot}$, and our range of late-flash–mixed models is $\leq 0.005\text{--}0.006M_{\odot}$ (the largest mass range found in the literature is $\leq 0.008M_{\odot}$; ref. 8).

Models including sedimentation. Guided by the evolutionary results, we built up sets of models, characterized by three main parameters: core mass M_{c} , envelope mass M_{env} , and initial hydrogen mass abundance in the envelope $X_{\text{env-in}}$, and followed their evolution, starting from the zero-age horizontal branch. As there are both helium-dominated and hydrogen-rich spectra among the blue hook stars^{17,33}, models need to include helium diffusion to correctly derive the ultraviolet magnitudes, which are strongly dependent on the T_{eff} values of the models. An increase in the flux is expected, due to the shift in T_{eff} that the star suffers when diffusion changes the helium–carbon dominated atmosphere into a hydrogen-rich one. The speed of diffusion, a byproduct of the residual turbulence in the outer envelope^{36,37}, must be calibrated to be compatible with the observed values of hydrogen abundance in the blue hook stars^{17,33}. In the models we calibrate the parameter M_{turb} . Mass loss can have a concomitant role, and, when included, M_{turb} must be larger to remain consistent with the observations. We show our choice among computations in Extended Data Fig. 2. The theoretical (Extended Data Fig. 2b) and observational (Extended Data Fig. 2a) Hertzsprung–Russell diagram evolution of the tracks having $X_{\text{env-in}}$ values ranging from 0.007 to 0.41 are plotted for two settings of diffusion efficiency. The F225W extension of the tracks increases from about 0.4 mag up to cover almost the full extension of the blue hook (histogram in Extended Data Fig. 2a), by increasing $X_{\text{env-in}}$ up to the maximum value. Extended Data Fig. 2c shows the time evolution of $X_{\text{env-in}}$ for the two choices of M_{turb} . A better representation of spectroscopic observations requires the use of the models with slower diffusion. Figure 2 shows the comparison of tracks and the location of simulated points (red dots) for a mixture of stars with different initial abundances, $X_{\text{env-in}} = 0.007$ (10%), 0.03 (45%) and 0.06 (45%). The T_{eff} location and the He/H observed abundances are instead less compatible if we

include a percentage of tracks having $X_{\text{env-in}} > 0.06$ (see also the simulation comparisons).

Assumptions for the rotating models. In the blue hook progeny of rotating stars the core mass at the helium-flash, $M_{\text{c}}(\omega)$, is larger than the core mass of late-flash–mixing models computed by standard stellar evolution. From models of evolution starting from solid-body rotation on the main sequence and angular-momentum conservation in shells¹⁸, we derive a parabolic expression for the helium-flash core-mass increase $\delta(M_{\text{c}})$ as a function of the angular velocity ω (in units of per second):

$$\delta(M_{\text{c}}/M_{\odot}) = 3.86 \times 10^5 \omega^2 \quad (1)$$

A self-consistent approach needs computation of rotating models starting from the zero-age main sequence (where we can safely assume a rotation rate acquired simply by angular momentum conservation) and including all possible mechanisms of transfer of angular momentum from the core to the envelope, plus the loss of momentum of the envelope due to magnetic wind. In the models available in the literature^{38,39}, the parameters necessary to model these mechanisms have been calibrated from the atmospheric abundance variations induced by the associated chemical mixing, but most of the sampled stars were slowly rotating from the beginning, given that the fraction of young main-sequence stars which are fast-rotating is quite small⁴⁰. The coexistence of a fast-rotating core and slow-rotating envelope will lead to strong differential rotation that can be responsible for strong chemical mixing along the red giant branch, compared to the mixing possible for initially slower-rotating stars, an outcome that may explain extreme abundance anomalies in some clusters^{24,41}, but it is not known how much these events may break down the inner fast core rotation. So we adopt the simple $M_{\text{c}}(\omega)$ relation¹⁸, which does not take into account the core–envelope interactions.

The most important parameter in our investigation is the main-sequence lifetime of these fast-rotating stars, which is longer than the lifetime of non-rotating stars of the same mass. If the age is fixed, the rotating evolving mass $M_{\text{ev}}(\omega)$ will be larger than the non-rotating mass. A fraction of the rotating stars evolves through the helium flash, otherwise there would be no blue hook, so we have to assume that its mass satisfies the relation:

$$M_{\text{ev}}(\omega) \geq M_{\text{c}}(\omega) + \Delta M_{\text{RG}}(\omega)$$

where $\Delta M_{\text{RG}}(\omega)$ is the total mass lost on the red giant branch. The flash luminosity in our models increases by $\delta L_{\text{flash}}/\delta M_{\text{c}} \approx 1.9 \times 10^4$ (where L and M_{c} are in solar units). Assuming Reimers’ mass loss rate, we expect an extra mass loss of 0.010–0.015 M_{\odot} for each extra 0.01 M_{\odot} increase in the core mass. If for example, we require $\delta M_{\text{c}} = 0.04 M_{\odot}$, corresponding to $\omega \approx 3.5 \times 10^{-4} \text{ s}^{-1}$, the evolving mass must be $\sim 0.08\text{--}0.10 M_{\odot}$ larger than the non-rotating mass, if it has to ignite the helium flash. If we assume a stronger dependence of mass loss on L than in Reimers’ mass loss rate, we need a larger $M_{\text{ev}}(\omega)$ value. Such an increase is compatible with the scarce existing estimates¹⁸ but a large computational effort is needed to solve the problem. A spread in the evolving mass $M_{\text{ev}}(\omega)$ is also useful to meet, at least for a range of ω , the strict requirements of late-flash–mixing conditions and to let the stars populate the blue hook: those outside the correct range will evolve into the helium-core white-dwarf remnants whose existence has been recently established¹⁵.

The core mass versus luminosity relation shows that the tip of the red giant branch is extended by $\delta \log[L/L_{\odot}] \approx 0.14$ (0.35 bolometric magnitudes) for a core mass increase of $\sim 0.04 M_{\odot}$. If the model developed here is correct, the brightest giants should belong to the helium-rich population, contrary to standard models (the core flash occurs at slightly smaller luminosities in the high- Y , non-rotating models).

From the theoretical to the observational plane. We use bolometric corrections for $Z = 0.0005$ and $[\alpha/\text{Fe}] = 0.2$ ($[\text{Fe}/\text{H}] = -1.74$) (where Z is the global metal mass fraction and $[\alpha/\text{Fe}]$ the ratio of α elements over iron with respect to the solar ratio) available for atmosphere models with standard hydrogen abundance⁴². A correction is necessary, because these spectra do not accurately represent the peculiar atmospheres of late-flash–mixed stars⁵, with strongly enhanced helium and carbon abundances at the surface. At a given T_{eff} we correct the helium-dominated model magnitudes, making them fainter by 0.057 mag in the F225W band, an estimate based on the comparison of fluxes in helium- and carbon-rich and in hydrogen-rich model atmospheres⁹. We keep this correction until the helium surface abundance remains above $Y = 0.9$, and then we switch to the direct table correlations. This choice contributes to the stretching (in magnitude extension) of the tracks with $0.03 \leq X_{\text{env-in}} \leq 0.06$.

Standard simulations. Synthetic models for the simulations shown in Fig. 1b follow standard guidelines⁴³. For the cases employing standard tracks, we assume that the red giant mass M_{RG} at a fixed age is a function of helium, at assumed metallicity. The mass on the horizontal branch is $M_{\text{RG}}(Y) - \Delta M$, where ΔM is the mass lost during the red giant phase. In our case, assuming $Y = 0.37$ and an age of 12 Gyr we derive $M_{\text{RG}} = 0.658 M_{\odot}$. The observed ratio of stars between

late-flash-mixed and ‘normal’ extreme horizontal branch stars is reproduced by assuming that ΔM has a Gaussian dispersion of $\sigma = 0.008$ around an average value $\Delta M = 0.19M_{\odot}$. With this choice, the resulting 60 extreme horizontal branch stars that did not suffer mixing are shown in Fig. 1b as red squares, and the 130 blue triangles are the late-flash-mixed stars. When we extract a mass smaller than the late-flash-mixed masses, the star is assumed to evolve into helium-core white dwarfs. With the chosen parameters we find 110 helium-core white dwarfs. Comparison is shown by histograms of counts as a function of colour and magnitude, given separately for extreme horizontal branch and blue hook stars. The bins used are 0.03 mag and 0.08 mag for colours and magnitudes respectively. The error bars are the result of individual count Poisson error, \sqrt{N} . The observational histograms are grey, the simulated ones are blue and red for the colour (scale on the right) and magnitude (scale on the top) distributions.

This modelling shows the inherent difficulty in building the blue hook: the mass range of extreme horizontal branch models (defined as the models at $T_{\text{eff}} > 20,000$ K) is a few hundredths of M_{\odot} (from $0.03M_{\odot}$ to $0.07M_{\odot}$, larger for a larger Y), while late-flash-mixing covers a mass range of only $\leq 0.008M_{\odot}$. The most recent modelling of horizontal branches of globular clusters shows that the mass-loss spread of each component of multiple populations must be very small, $< 0.01M_{\odot}$. If the blue hook standard models are truly describing the blue hook stars in globular clusters, the mass lost by the red giant progenitor must be finely tuned, so that it covers precisely this tiny mass range. If the blue hook is made up by more than one of the globular cluster populations (groups of stars with different helium and metal contents) such a fine tuning must have been achieved twice. This is a key issue in interpreting the simulations.

First we assume that all the extreme horizontal branch stars have $Y = 0.37$. We use this Y value because it is consistent with the requirements of the asymptotic giant branch model for the formation of the extreme populations in clusters⁴⁴. We use different sets of flash-mixing models, with $X_{\text{env-in}} = 0.007, 0.03, 0.06, 0.21$ and 0.41 . The simulations with $X_{\text{env-in}} \leq 0.06$ are unsatisfactory, because they do not cover the whole blue hook extension (Fig. 1b). Including a fraction of stars with $X_{\text{env-in}} = 0.21$, the magnitude range is marginally better covered, but the colour agreement is worse (Extended Data Fig. 4b). The magnitude range is not extended much because the track having $X_{\text{env-in}} = 0.06$ already reaches high atmospheric hydrogen content along the evolution including hydrogen versus helium diffusion (Extended Data Fig. 2c). Including even stars with $X_{\text{env-in}} = 0.41$, these latter stars merge with the extreme horizontal branch (Extended Data Fig. 4c).

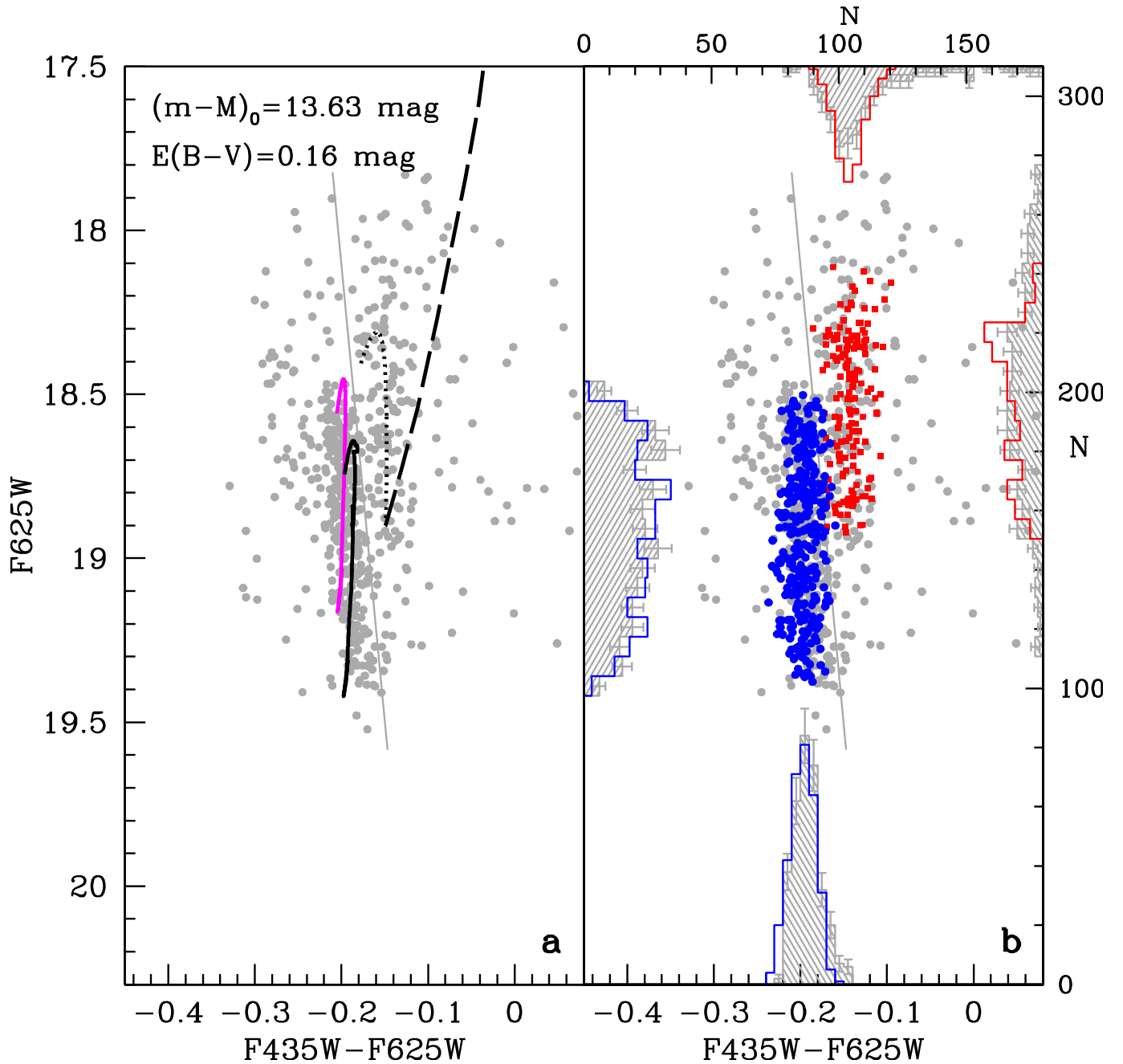
Extended Data Fig. 3 shows the simulation according to which the blue hook is made up by superposition of two different late-flash-mixing populations, with $Y = 0.25$ and $Y = 0.37$, respectively⁴⁵. ΔM and σ values used are reported in the figure. The mass losses necessary to obtain masses in the late-flash-mixing range differ by about $0.13M_{\odot}$ for the two different Y groups. It is already very difficult to accept that the mass-loss of a unique population is so well constrained that the blue hook is populated at all, but this double constraint is an even more extreme assumption. In spite of this, the full extension of the hook is not yet accounted for.

Coupled dynamical and evolutionary simulations. To estimate the distribution of rotation rates we use a simple semi-analytical model to calculate the distribution of times, T_{enc100} , needed for second-generation star-disk systems to have a close encounter with another star at a distance $d < 100$ astronomical units (AU) (the estimated dimension of the protostellar disk) and assume one—or more—such collisions to be able to break the magnetic disk-locking (in this case T_{enc100} is the detachment time). To calculate T_{enc100} we have followed the orbits of 50,000 particles distributed as a King model with a concentration c of about 1.7 and integrated the collision rate (see equation 7.194 in ref. 46) along the orbit to

estimate T_{enc100} . The system we are modelling is meant to represent the dense second-generation subcluster. For our calculation we assume its total mass to be equal to $10^{5.5}M_{\odot}$ and its half-mass radius about one parsec. From the distribution of T_{enc100} (Extended Data Fig. 5a) we directly derive the rotation rates assuming the conservation of angular momentum from Fig. 3, and the corresponding increase in the core mass from equation (1) in Methods (Extended Data Fig. 5c). The initial periods P_{in} at detachment from the disk are randomly extracted in the range $3 < P$ (days) < 12 , using a normal distribution centred at $P = 6$ days, with standard deviation $\sigma = 2$ days. The simulation results for the distribution of the stellar rotation periods at detachment from the disk are given in Extended Data Fig. 5b. The distribution is not Gaussian, as it also depends on the limitations we have imposed on the maximum δM_c . For the photometric simulation, we then assume the dispersion in core masses inferred by the described simulation (Extended Data Fig. 5c). We are implicitly assuming that the increase in initial evolving mass due to rotation is enough to accommodate the increase both in core mass and mass loss in such a way that the stars are able to arrive at late-flash-mixing ignition for rotation rates $\omega > 10^{-6} \text{ s}^{-1}$. We also impose the condition that a tail of slower rotating stars populate the extreme horizontal branch of non-flash-mixed $Y = 0.37$ models.

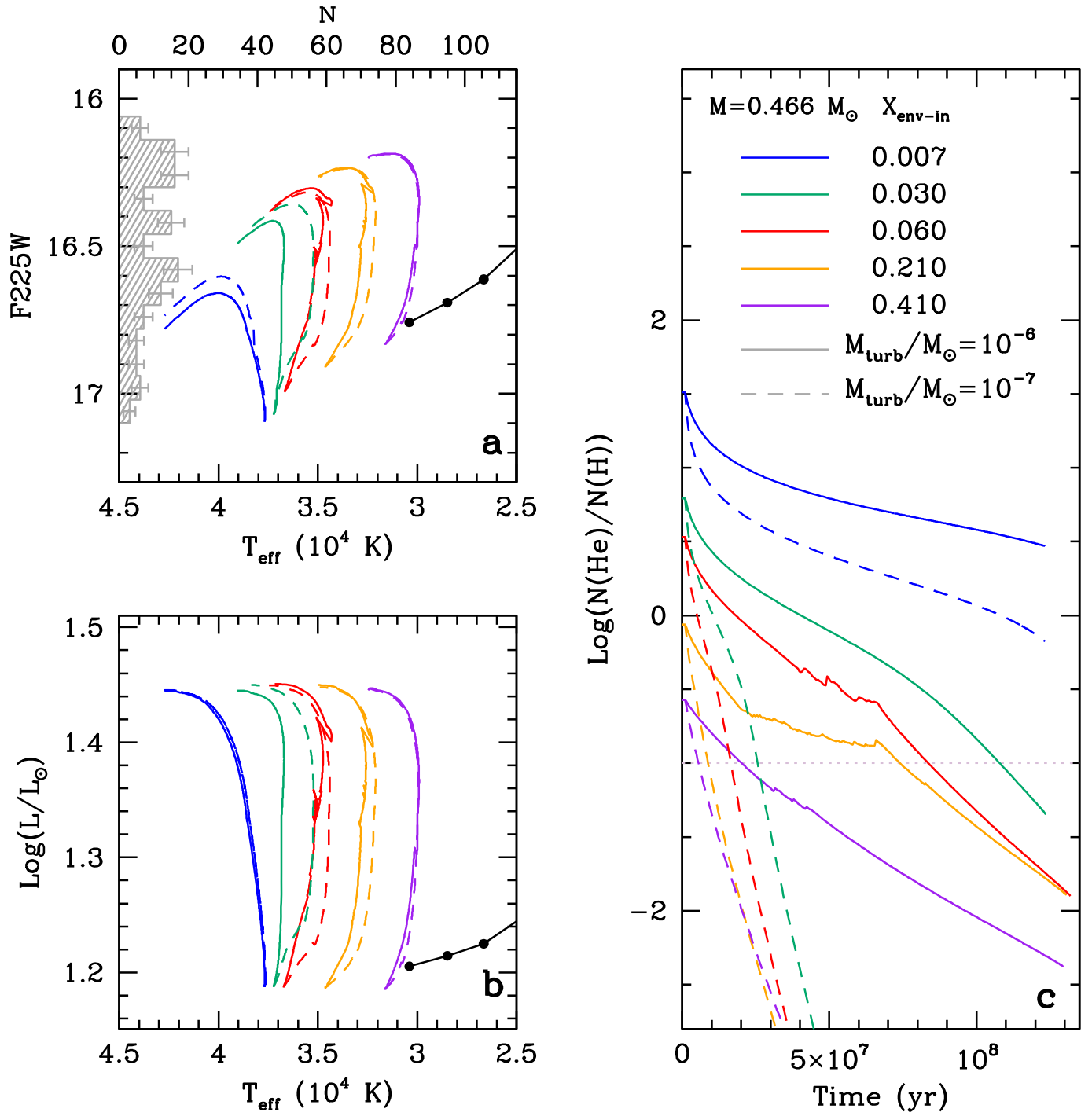
Code availability. For the stellar evolution calculations we used ATON (not publicly available). For the N -body simulations we used Starlab (publicly available at <http://www.sns.ias.edu/~starlab/>). For the synthetic populations, we used a program (HBSINT) written explicitly for this work (not publicly available).

31. Cardelli, J. A., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345**, 245–256 (1989).
32. O'Donnell, J. E. R_{UV} -dependent optical and near-ultraviolet extinction. *Astrophys. J.* **422**, 158–163 (1994).
33. Moehler, S. *et al.* The hot horizontal-branch stars in ω Centauri. *Astron. Astrophys.* **526**, A136 (2011).
34. Ventura, P., D'Antona, F. & Mazzitelli, I. The ATON 3.1 stellar evolutionary code. A version for asteroseismology. *Astrophys. Space Sci.* **316**, 93–98 (2008).
35. Castellani, M. & Castellani, V. Mass loss in globular cluster red giants—an evolutionary investigation. *Astrophys. J.* **407**, 649–656 (1993).
36. Michaud, G., Richer, J. & Richard, O. Horizontal branch evolution and atomic diffusion. *Astrophys. J.* **670**, 1178–1187 (2007).
37. Michaud, G., Richer, J. & Richard, O. Abundance anomalies in horizontal branch stars and atomic diffusion. *Astrophys. J.* **675**, 1223–1232 (2008).
38. Gallet, F. & Bouvier, J. Improved angular momentum evolution model for solar-like stars. *Astron. Astrophys.* **556**, A36 (2013).
39. Palacios, A., Talon, S., Charbonnel, C. & Forestini, M. Rotational mixing in low-mass stars. I. Effect of the μ -gradients in main sequence and subgiant Pop I stars. *Astron. Astrophys.* **399**, 603–616 (2003).
40. Spada, F., Lanzafame, A. C., Lanza, A. F., Messina, S. & Collier Cameron, A. Modelling the rotational evolution of solar-like stars: the rotational coupling time-scale. *Mon. Not. R. Astron. Soc.* **416**, 447–456 (2011).
41. Charbonnel, C. & Talon, S. Influence of gravity waves on the internal rotation and Li abundance of solar-type stars. *Science* **309**, 2189–2191 (2005).
42. Castelli, F. & Kurucz, R. L. New grids of ATLAS9 model atmospheres. Preprint at <http://arxiv.org/abs/astro-ph/0405087> (2004).
43. D'Antona, F. & Caloi, V. The fraction of second generation stars in globular clusters from the analysis of the horizontal branch. *Mon. Not. R. Astron. Soc.* **390**, 693–705 (2008).
44. Ventura, P. & D'Antona, F. Hot bottom burning in the envelope of super asymptotic giant branch stars. *Mon. Not. R. Astron. Soc.* **410**, 2760–2766 (2011).
45. Cassisi, S. *et al.* Hot horizontal branch stars in ω Centauri: clues about their origin from the cluster color magnitude diagram. *Astrophys. J.* **702**, 1530–1535 (2009).
46. Binney, J. & Tremaine, S. *Galactic Dynamics* 2nd edn (Princeton Univ. Press, 2008).



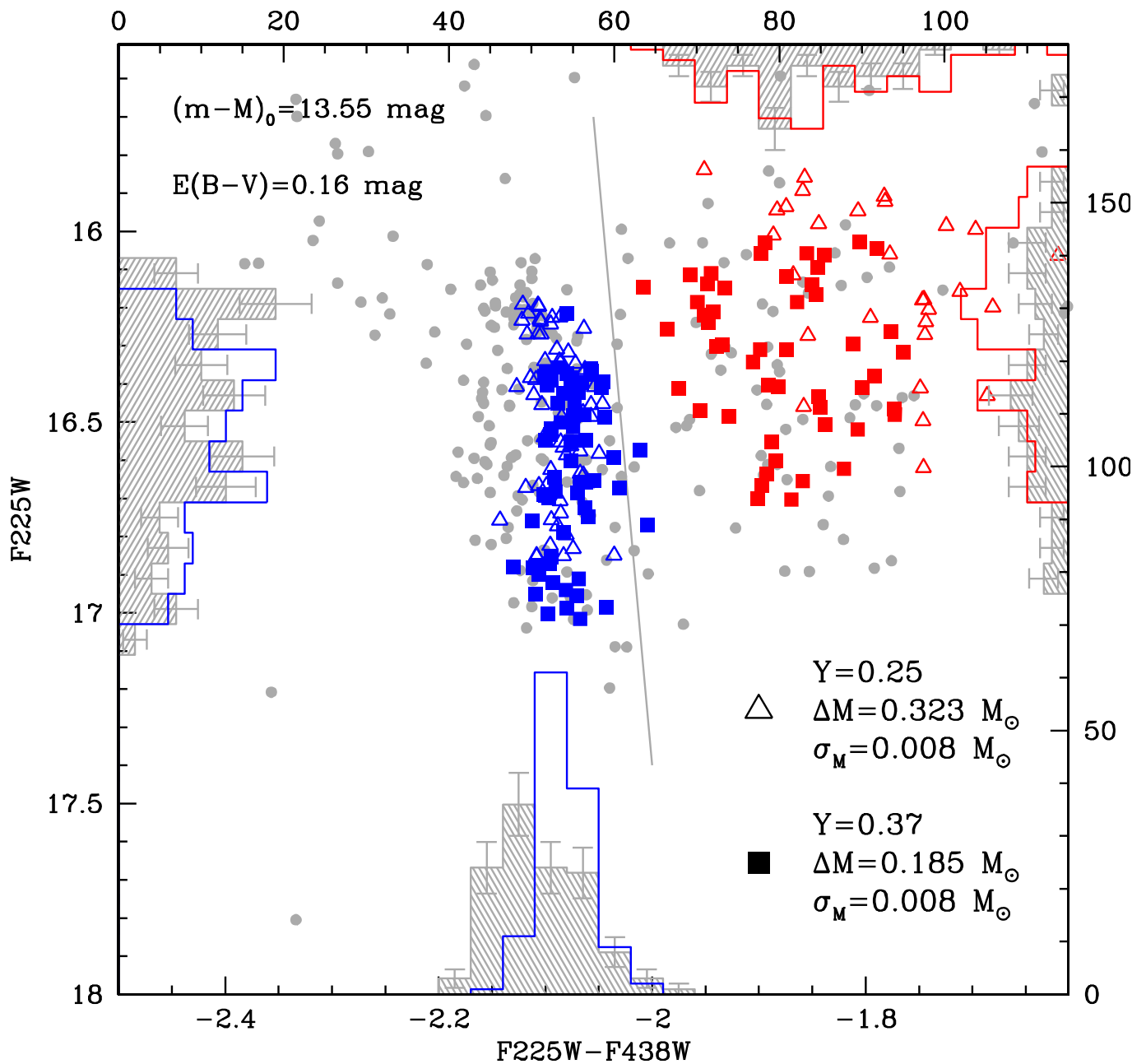
Extended Data Figure 1 | Comparison simulation versus optical data. The observed data¹⁰ are shown as grey dots. The diagonal line suggests the division between blue hook and cooler extreme horizontal branch stars. **a**, Tracks shown are for $M = 0.466M_{\odot}$, $Y = 0.37$ and evolutionary $M_c = 0.463M_{\odot}$ (black solid line), plus the corresponding track in which M_c is increased by

$0.04M_{\odot}$ (magenta solid line). The zero-age horizontal branch for $Y = 0.37$ (dashed black) and the $0.471M_{\odot}$ track (dotted black) are also shown. **b**, Result of the coupled dynamical-evolutionary simulation (detailed explanation in Fig. 1).



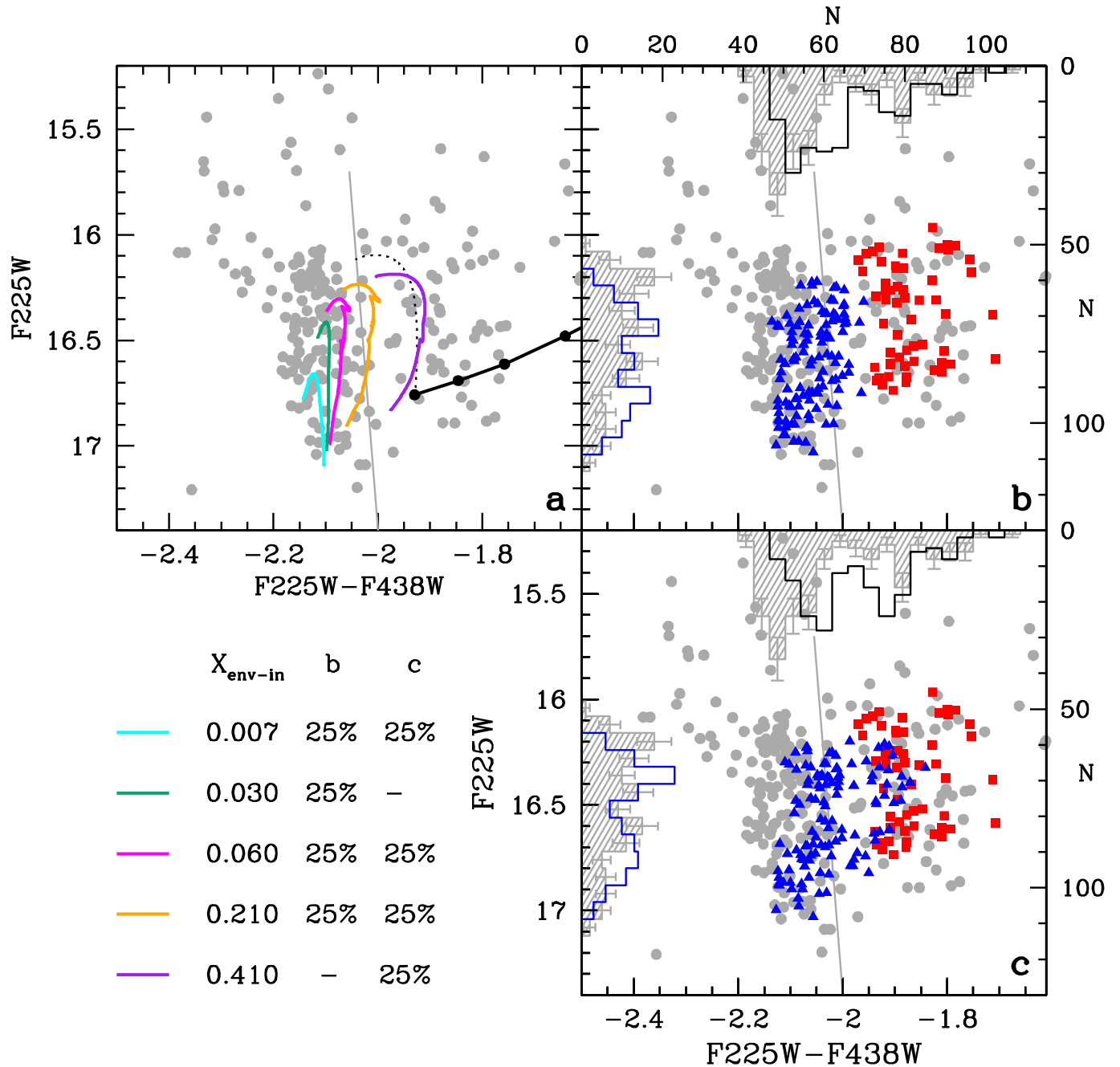
Extended Data Figure 2 | Models with helium versus hydrogen diffusion. **a** and **b** show a comparison of blue hook tracks from $X_{\text{env-in}} = 0.007$ to $X_{\text{env-in}} = 0.41$, from left to right, for $M_{\text{turb}} = 10^{-6} M_{\odot}$ (full lines) and $10^{-7} M_{\odot}$ (dashed lines). The bolometric luminosity (**b**) and the F225W magnitude (**a**) are shown as function of T_{eff} . The hot side of the zero-age horizontal branch

for $Y = 0.37$ is shown (solid black line with dots). **c**, Evolution with time of the surface $N(\text{He})/N(\text{H})$ due to diffusion. The solar ratio is shown as a horizontal grey dotted line. The observations of blue hook stars show that the slower diffusion is a better description (Fig. 2).



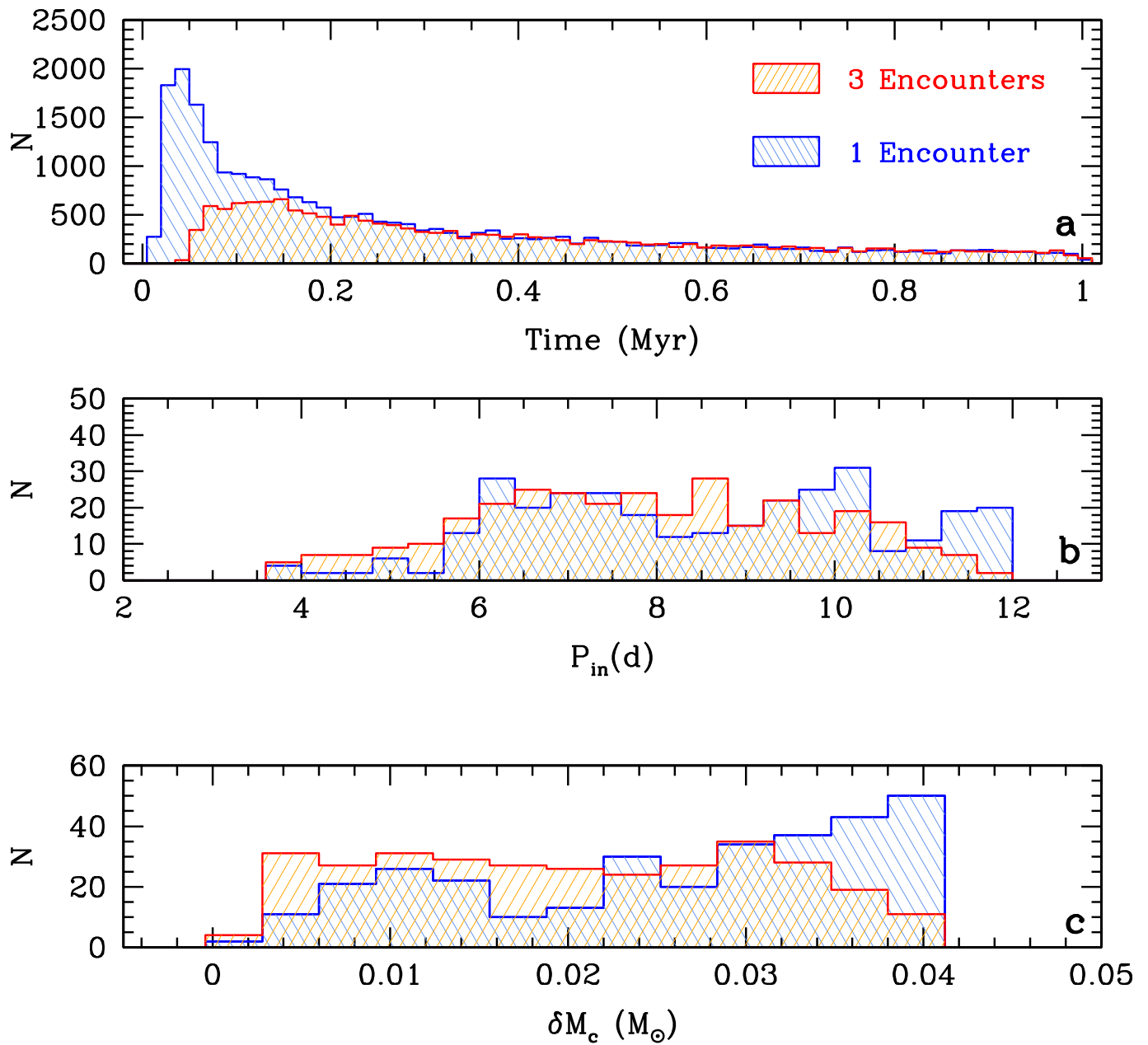
Extended Data Figure 3 | Simulation versus data, including first- and second-generation stars. As in Fig. 1 and Extended Data Fig. 1, but the ultraviolet data are compared with a simulation assuming that 50% of blue hook

stars are the progeny of the $Y = 0.37$ population (squares), and 50% are progeny of the $Y = 0.25$ population (triangles). The error bars are the result of individual 1σ Poisson error on the stellar counts.



Extended Data Figure 4 | Tracks and simulation versus data including models having $X_{\text{env-in}} > 0.06$. As in Fig. 1. a, Tracks with different $X_{\text{env-in}}$ values and the end of the zero-age horizontal branch (black line with dots) with the location of the minimum non-mixed track (dotted black line) for $Y = 0.37$.

b, Simulation including $X_{\text{env-in}}$ up to 0.21, for the listed percentages of stars. c, Simulation also including stars with $X_{\text{env-in}} = 0.41$. In the latter simulation, the main aim is to show that the most hydrogen-rich stars are located on the extreme horizontal branch, and not on the blue hook.



Extended Data Figure 5 | The dynamical model. Two different cases of dynamical interactions are compared: either we make the hypothesis that the magnetic coupling of the disk is destroyed by one encounter with another star, at a distance entering the accretion disk (<100 AU), or that three such

encounters are necessary. **a**, A histogram of encounters versus time of detachment. **b**, Initial rotation period distribution employed in the two cases. **c**, Resulting distribution of the core mass increase, which is an initial parameter in the photometric simulation.

Extended Data Table 1 | Range in magnitude F225W for $M = 0.466M_{\odot}$, $M_c = 0.463M_{\odot}$

$X_{\text{env-in}}$	$F225W$ ZAHB	$F225W$ Peak	$F225W$ TAHB	$\delta F225W$ (track)	$\delta F225W$ total	$(F225W - F438W)_0$ ZAHB	T_{eff} ZAHB
0.007	+2.29	+1.86	+1.97	0.43	0.43	−2.734	37660
0.030	+2.21	+1.61	+1.69	0.60	0.68	−2.731	37200
0.060	+2.19	+1.50	+1.56	0.69	0.79	−2.726	36700
0.210	+2.10	+1.43	+1.46	0.67	0.83	−2.693	34500
0.410	+2.03	+1.38	+1.39	0.65	0.90	−2.603	31600
$\Delta M_{\text{core}} = +0.04M_{\odot}$							
0.007	+1.96	+1.58	+1.76	0.38	0.71	−2.700	39400
0.030	+1.93	+1.38	+1.71	0.55	0.91	−2.697	38900
0.060	+1.91	+1.27	+1.35	0.64	1.02	−2.693	38400

For each initial envelope starting hydrogen abundance $X_{\text{env-in}}$ (column 1), we give F225W band values of late-flash–mixed models at the zero-age horizontal branch ZAHB (column 2), at the maximum luminosity (column 3) and at the terminal point (column 4) (defined as the model in which the core helium abundance becomes zero), and the magnitude interval spanned by each model (column 5) and by a mixture (column 6). The mass of the model is $0.466M_{\odot}$, with an envelope mass of $0.003M_{\odot}$, so the model core mass is $0.463M_{\odot}$.

Extended Data Table 2 | Data at 12 Gyr for models of $Z = 0.0005$, $[\alpha/\text{Fe}] = 0.2$

Y	M_{tip}	$M_{HB_{\text{min}}}$	M_{core}	M_{LFM}	$M_{\text{core}}(LFM)$	Δ_{MLF}
0.25	0.812	0.500	0.496	0.489	0.486	0.323
0.37	0.657	0.471	0.469	0.466	0.463	0.191
0.40	0.620	0.466	0.465	0.462	0.459	0.158

We list some important quantities from our isochrones of 12 Gyr, metallicity $Z = 0.0005$ and $[\alpha/\text{Fe}] = 0.2$. Column 1 shows the helium mass fraction Y ; column 2 shows the evolving mass at the tip of the red giant branch (no mass loss); column 3 shows the minimum 'standard' horizontal branch mass; column 4 shows the core mass in M_{\odot} , when the helium flash is ignited at the tip of the red giant branch; column 5 shows the mass for which we have late-flash mixing; column 6 shows the core mass for the late-flash-mixed model; and column 7 shows the mass loss needed to achieve late-flash mixing. All masses are in solar units.

Laboratory confirmation of C_{60}^+ as the carrier of two diffuse interstellar bands

E. K. Campbell¹, M. Holz¹, D. Gerlich² & J. P. Maier¹

The diffuse interstellar bands are absorption lines seen towards reddened stars¹. None of the molecules responsible for these bands have been conclusively identified². Two bands at 9,632 ångströms and 9,577 ångströms were reported in 1994, and were suggested to arise from C_{60}^+ molecules (ref. 3), on the basis of the proximity of these wavelengths to the absorption bands of C_{60}^+ measured in a neon matrix⁴. Confirmation of this assignment requires the gas-phase spectrum of C_{60}^+ . Here we report laboratory spectroscopy of C_{60}^+ in the gas phase, cooled to 5.8 kelvin. The absorption spectrum has maxima at $9,632.7 \pm 0.1$ ångströms and $9,577.5 \pm 0.1$ ångströms, and the full widths at half-maximum of these bands are 2.2 ± 0.2 ångströms and 2.5 ± 0.2 ångströms, respectively. We conclude that we have positively identified the diffuse interstellar bands at 9,632 ångströms and 9,577 ångströms as arising from C_{60}^+ in the interstellar medium.

Shortly after the discovery of C_{60} (ref. 5), the question of its relevance to the diffuse interstellar bands (DIBs) was raised⁶. It soon became apparent that neutral C_{60} does not have the appropriate absorptions in the DIB range⁷. However, owing to its low ionization potential, it was pointed out⁸ that C_{60} would be present mainly as C_{60}^+ in the diffuse interstellar clouds and hence could be a DIB carrier. Charge transfer complexes of $(C_{60}-M)^+$, where M is any cosmically abundant atom, have also been considered and proposed as further candidates⁸. With the discovery of C_{60} in planetary⁹ and reflection nebula¹⁰, interest in C_{60}^+ has been renewed.

We built an apparatus to measure the electronic spectra of cations that are of astrophysical interest at temperatures that are typical of the interstellar medium¹¹. The central part of the instrument is a cryogenic 22-pole radio-frequency trap¹², in which ions are confined and undergo collisions with cold (5 K) and very dense ($4 \times 10^{15} \text{ cm}^{-3}$) helium. In the case of smaller polyatomic cations, we were able to show spectroscopi-

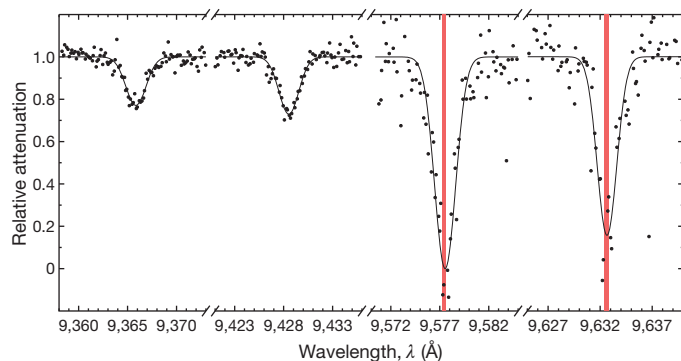


Figure 1 | Gas-phase laboratory spectra of C_{60}^+ at 5.8 K. The spectra were recorded by monitoring the depletion on the C_{60}^+ -He mass channel but, as described in the text, the effect of the helium is so small (less than 0.2 Å) that they can be considered as C_{60}^+ spectra for comparison with astronomical data. Gaussian fits to the experimental data (circles) are represented by the solid black lines. The intensities of the bands have been scaled by the measured relative absorption cross-sections. The fit parameters are given in Table 1. The vertical red lines are the rest wavelengths, $9,577.4 \pm 0.2$ Å and $9,632.6 \pm 0.2$ Å, of two reported DIBs¹⁷; their widths represent the uncertainty.

cally that there are a sufficient number of collisions to ensure that the rotational and vibrational temperatures are equilibrated¹³. Under these conditions it became possible to synthesize *in situ* helium complexes of mass-selected ions via ternary association¹⁴.

These experimental advances allowed us to obtain the electronic spectrum of C_{60}^+ -He by one-photon excitation followed by the loss of the helium atom. Such methods have been used for decades in spectroscopy of ions, mainly in the infrared¹⁵ using rare gases, but also in the visible with helium attachment¹⁶. Owing to the low binding energies of weakly bound complexes such as C_{60}^+ -He, temperatures below 8 K are required.

The measured spectra of C_{60}^+ , presented in Fig. 1, are in very close agreement with the two DIBs at 9,632 Å and 9,577 Å. As is expected from the weak interaction of the helium, measurement of the corresponding absorption for C_{60}^+ -He₂ (Fig. 2) indicates that the shift on the electronic transition is less than 0.2 Å.

The best fits to our data with a single function are obtained using a Gaussian function, rather than a Lorentzian. The band maxima are at $9,632.7 \pm 0.1$ Å and $9,577.5 \pm 0.1$ Å. The rest wavelengths for the DIBs are $9,632.6 \pm 0.2$ Å and $9,577.4 \pm 0.2$ Å after correction for stellar lines¹⁷. The extracted wavelengths towards different stars are nearly all within 1 Å of those given above^{17–19}. This agreement with our laboratory data is remarkable, especially considering the corrections needed in the determination of the DIBs. In astronomical measurements, small differences in the wavelength of the band maximum, band full width at half-maximum (FWHM), and relative equivalent widths arise from the environmental conditions within the interstellar clouds, the number of clouds sampled in the line of sight and the accuracy of the telluric corrections.

Further support for the assignment of C_{60}^+ as a DIB carrier comes from the widths of the two bands at 9,632.7 Å and 9,577.5 Å. The FWHM value of the interstellar bands that was reported following observations towards HD183143 is 2.85 ± 0.2 Å (ref. 18). Our laboratory results are 2.2 ± 0.2 Å and 2.5 ± 0.2 Å for the maxima at $9,632.7 \pm 0.1$ and $9,577.5 \pm 0.1$ Å, respectively. The FWHM of the C_{60}^+ rotational profile at 5.8 K is about 1 Å (ref. 20). The lines are therefore broadened by internal conversion, indicating a lifetime of 2 ps of the excited electronic state.

The somewhat larger FWHM of the DIBs is expected because of the higher rotational temperature in the diffuse clouds. Temperatures of non-polar molecules are higher than the 5.8 K in our measurements; for example, H_3^+ (ref. 21) and C_3 (ref. 22) lie in the 30–80 K range.

Table 1 | Gas-phase band maxima, widths and absorption cross-sections

λ (Å)	FWHM (Å)	σ_{rel}
$9,632.7 \pm 0.1$	2.2 ± 0.2	0.8
$9,577.5 \pm 0.1$	2.5 ± 0.2	1
$9,428.5 \pm 0.1$	2.4 ± 0.1	0.3
$9,365.9 \pm 0.1$	2.4 ± 0.1	0.2

The wavelengths, λ , band FWHMs and standard deviations are determined from fits to the C_{60}^+ -He fragmentation spectra using a single Gaussian function. The experimentally determined relative absorption cross-sections, σ_{rel} , are a measure of the relative intensities, with an estimated uncertainty of about 20%.

¹Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland. ²Department of Physics, Technische Universität, 09107 Chemnitz, Germany.

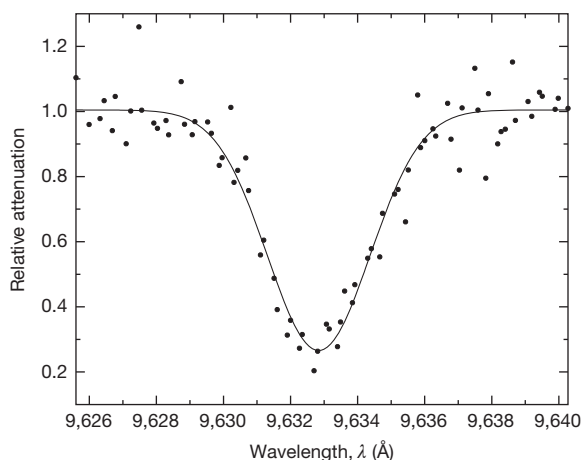


Figure 2 | C_{60}^+ - He_2 spectrum. This spectrum was recorded by monitoring the depletion on the C_{60}^+ - He_2 mass channel. A Gaussian fit to the experimental data (circles) is represented by the solid line. The fit yields a band maximum at $9,632.7 \pm 0.1$ Å and a FWHM of 3.6 ± 0.2 Å.

The relative cross-sections of the $9,632.7 \pm 0.1$ Å and $9,577.5 \pm 0.1$ Å laboratory absorptions were determined (Methods, Fig. 3) to be about equal, in good agreement with the intensities of the DIBs. The equivalent widths of the 9,632 Å and 9,577 Å DIBs vary somewhat in the literature^{17–19}, and the 9,632 Å DIB is partially blended with a Mg(II) line. The consensus seems to be that the equivalent widths of the DIBs are comparable, or perhaps that the equivalent width of the 9,577 Å DIB is a bit larger.

In the study of the absorption spectrum in the neon matrix⁴, it was not clear what the two transitions close to the DIBs were caused by. The current experiments clarify this: irradiation at either 9,632.7 Å or 9,577.5 Å leads to near-complete attenuation in the number of complexes, as shown in Fig. 3. This result indicates that the two transitions arise from a single structural isomer. The implications of our findings are that both transitions originate from the lowest vibrational level of the ground electronic state, $^2A_{1u}$ in D_{5d} symmetry, and result in two excited states separated by 55.2 Å. These are either two electronic states, suggested on the basis of magnetic circular dichroism

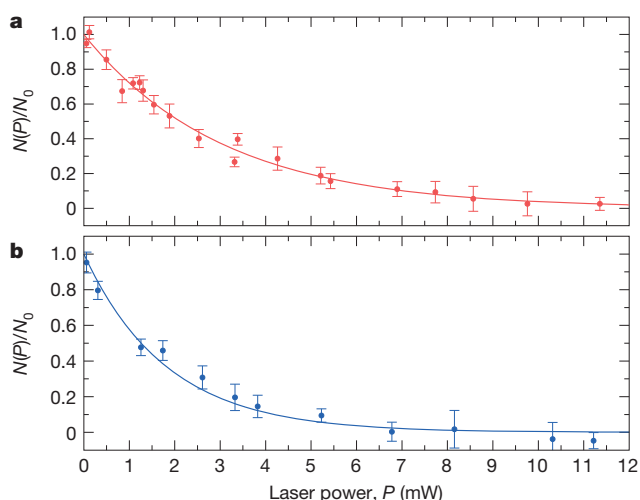


Figure 3 | Relative cross-section measurement. a, b, Depletion on the C_{60}^+ - He mass channel as a function of laser power at 9,632 Å (a) and 9,577 Å (b). $N(P)$ is the number of C_{60}^+ - He complexes (as a function of laser power, P); $N_0 = N(0)$. Experimental data (circles, ± 1 s.d.) were corrected for the number of background ions appearing at a mass-to-charge ratio of 724 AMU/e (for example, $^{13}C_4^{12}C_{56}$). Exponential fits (solid lines, see Methods) provide information on the absorption cross-sections and indicate that all trapped ions interact with the laser.

measurements in argon matrices, or the two spin-orbit components of the upper state, 2E_g in D_{5d} (refs 23, 24).

According to the absorption spectrum in the neon matrix⁴, the next two bands on the short-wavelength side of the 9,577 Å peak are weaker by a factor of 4–5 in intensity. A search for DIBs at 9,366 Å and 9,419 Å was reported^{17,19}. In this region, we observed bands in the gas phase at $9,365.9 \pm 0.1$ Å and $9,428.5 \pm 0.1$ Å, as shown in Fig. 1. The lack of a detected DIB at 9,366 Å was assumed to indicate an upper limit of 16% of the 9,577 Å band intensity¹⁷. This conclusion is consistent with the relative cross-section of about 20% that we determined. The other band was not found, despite searching for it in interstellar clouds at 9,419 Å. However, this absorption is at $9,428.5 \pm 0.1$ Å in the gas phase, according to our measurement. A ‘depression’ at 9,428 Å was reported¹⁷. Astronomical measurements in this region are difficult because of stellar lines and telluric absorptions from water vapour.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 April; accepted 11 May 2015.

- Herbig, G. H. The diffuse interstellar bands. *Annu. Rev. Astron. Astrophys.* **33**, 19–73 (1995).
- Snow, T. P. & McCall, B. J. Diffuse atomic and molecular clouds. *Annu. Rev. Astron. Astrophys.* **44**, 367–414 (2006).
- Foing, B. H. & Ehrenfreund, P. Detection of two interstellar absorption bands coincident with spectral features of C_{60}^+ . *Nature* **369**, 296–298 (1994).
- Fulara, J., Jakobi, M. & Maier, J. P. Electronic and infrared spectra of C_{60}^+ and C_{60}^- in neon and argon matrices. *Chem. Phys. Lett.* **211**, 227–234 (1993).
- Kroto, H. W., Heath, J. R., O’Brien, S. C., Curl, R. F. & Smalley, R. E. C_{60} : Buckminsterfullerene. *Nature* **318**, 162–163 (1985).
- Kroto, H. W. Space, stars, C_{60} , and soot. *Science* **242**, 1139–1145 (1988).
- Herbig, G. H. The search for interstellar C_{60} . *Astrophys. J.* **542**, 334–343 (2000).
- Kroto, H. W. & Jura, M. Circumstellar and interstellar fullerenes and their analogues. *Astron. Astrophys.* **263**, 275–280 (1992).
- Cami, J., Bernard-Salas, J., Peeters, E. & Malek, S. E. Detection of C_{60} and C_{70} in a young planetary nebula. *Science* **329**, 1180–1182 (2010).
- Sellgren, K. et al. C_{60} in reflection nebulae. *Astrophys. J.* **722**, L54–L57 (2010).
- Chakrabarty, S. et al. A novel method to measure electronic spectra of cold molecular ions. *J. Phys. Chem. Lett.* **4**, 4051–4054 (2013).
- Gerlich, D. Ion-neutral collisions in a 22-pole trap at very low energies. *Phys. Scr.* **T59**, 256–263 (1995).
- Chakrabarty, S., Rice, C. A., Mazzotti, F. J., Dietsche, R. & Maier, J. P. Electronic absorption spectrum of triacetylene cation for astronomical considerations. *J. Phys. Chem. A* **117**, 9574–9577 (2013).
- Jašík, J., Zábka, J., Roithová, J. & Gerlich, D. Infrared spectroscopy of trapped molecular dications below 4 K. *Int. J. Mass Spectrom.* **354–355**, 204–210 (2013).
- Duncan, M. A. Infrared laser spectroscopy of mass-selected carbocations. *J. Phys. Chem. A* **117**, 11477–11491 (2012).
- Bieske, E. J., Soliva, A. M., Friedmann, A. & Maier, J. P. Electronic spectra of $N_2^+-(He)_n$ ($n = 1, 2, 3$). *J. Chem. Phys.* **96**, 28–34 (1992).
- Jenniskens, P., Mulas, G., Porceddu, I. & Benvenuti, P. Diffuse interstellar bands near 9600 Å: not due to C_{60}^+ yet. *Astron. Astrophys.* **327**, 337–341 (1997).
- Foing, B. H. & Ehrenfreund, P. New evidences for interstellar C_{60}^+ . *Astron. Astrophys.* **317**, L59–L62 (1997).
- Galazutdinov, G. A., Krelowski, J., Musaev, F. A., Ehrenfreund, P. & Foing, B. H. On the identification of the C_{60}^+ interstellar features. *Mon. Not. R. Astron. Soc.* **317**, 750–758 (2000).
- Edwards, S. A. & Leach, S. Simulated rotational band contours of C_{60} and their comparison with some of the diffuse interstellar bands. *Astron. Astrophys.* **272**, 533–540 (1993).
- Indriolo, N., Geballe, T. R., Oka, T. & McCall, B. J. H_3^+ in diffuse interstellar clouds: a tracer for the cosmic-ray ionization rate. *Astrophys. J.* **671**, 1736–1747 (2007).
- Maier, J. P., Lakin, N. M., Walker, G. A. H. & Bohlander, D. A. Detection of C_3 in diffuse interstellar clouds. *Astrophys. J.* **553**, 267–273 (2001).
- Gasyňa, Z., Andrews, L. & Schatz, P. N. Near-infrared absorption spectra of C_{60} radical cations and anions prepared simultaneously in solid argon. *J. Phys. Chem.* **96**, 1525–1527 (1992).
- Langford, V. S. & Williamson, B. E. Magnetic circular dichroism of C_{60}^+ and C_{60}^- radicals in argon matrices. *J. Phys. Chem. A* **103**, 6533–6539 (1999).

Acknowledgements This work was financially supported by the European Research Council (ERC-AdG-ElecSpecIons: 246998)

Author Contributions E.K.C., M.H. and D.G. recorded and analysed the experimental data. J.P.M. and E.K.C. wrote the paper with input from all authors. All authors discussed the results and commented on the manuscript. J.P.M. initiated and led the project.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.P.M. (j.p.maier@unibas.ch).

METHODS

The use of cryogenic ion traps for astrophysics and spectroscopy is well documented in the literature^{11–14}.

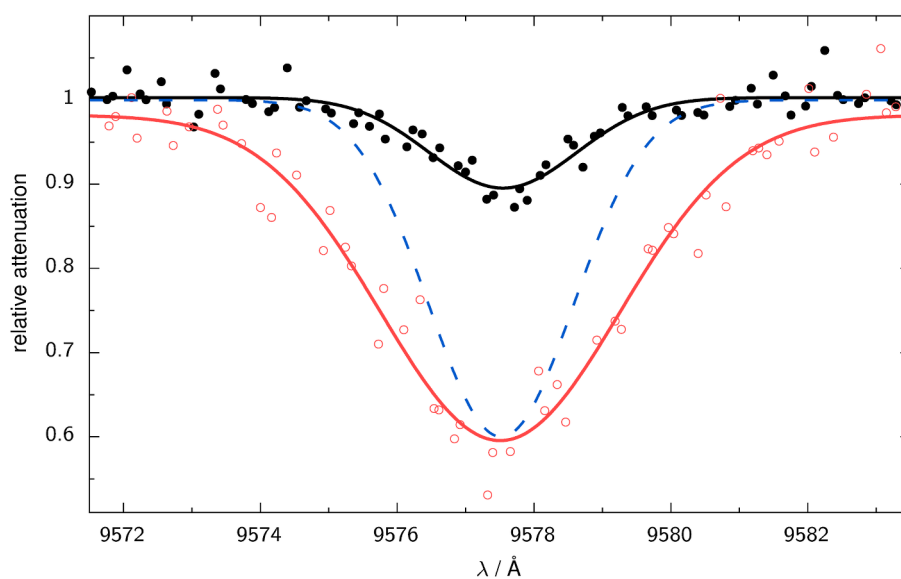
Experimental procedure. C_{60}^+ was produced using 50 eV electron impact of the neutral gas at 10^{-4} mbar. After passing through a quadrupole mass filter, operated in the transmission mode, the internally hot ions are deflected by 90 degrees using an electrostatic quadrupole bender and injected into a 22-pole radio frequency (amplitude, $V_0 = 160$ V; frequency, $f = 4.85$ MHz) ion trap. The use of the bender allowed for the separation of neutrals and ions. The trap is mounted onto the second stage of a closed-cycle cryostat (Sumitomo, RDK-205E). Trapping is achieved by pulsing the potential of the entrance and exit electrodes; the trap contents are analysed using a quadrupole mass spectrometer and a Daly detector. Experiments are performed at a repetition rate of 1 Hz.

The internal degrees of freedom of the ions are cooled via inelastic collisions with helium buffer gas, which is in equilibrium with the temperature of the trap walls (nominal temperature, $T_{\text{nom}} = 5$ K). The C_{60}^+ -He ions are rotationally cold because the internal temperature of the ion (T_{rot}) is given by the mass-weighted average of the translational temperature of the ions and the buffer gas, $T_{\text{rot}} = (m_1 T_1 + m_2 T_2)/(m_1 + m_2)$, where m_1 , T_1 and m_2 , T_2 are the mass and translational temperature of the ions and buffer gas, respectively. For heavy C_{60}^+ -He stored in cold helium, a translational temperature of 150 K for the ions still leads to a rotational temperature $T_{\text{rot}} = 5.8$ K.

About 10^5 C_{60}^+ cm^{-3} per filling are loaded into the trap by lowering the potential of the entrance electrode for 200 ms. Here they interact with high-number-density helium buffer gas, $[\text{He}] = 4 \times 10^{15} \text{ cm}^{-3}$, for 500 ms. The helium is introduced by resonantly exciting a piezo valve with amplitude 4.5 V at a frequency of 3.8 kHz. In the trap, C_{60}^+ -He complexes are formed via ternary association. Owing to extremely slow cooling of all vibrational modes to their ground states, attachment of helium to primary ions is very inefficient. However, a few per cent are sufficient, especially because a large fraction can be

fragmented. After pumping out the gas for 100 ms, the ion cloud is exposed to continuous-wave radiation that is produced from a homebuilt diode laser (up to 100 mW, 12 MHz bandwidth), which is gated with a mechanical shutter and open for a period of 300 ms. To avoid power broadening, only 2 mW were used for recording the two bands at $9,632.7 \pm 0.1$ Å and $9,577.5 \pm 0.1$ Å. For the other two, at 9,428.5 Å and 9,365.9 Å, 20–30 mW were required to obtain an attenuation of 20%. The potential of the exit electrode is lowered, and the contents extracted and analysed 150 ms after irradiation. The resulting laser-induced attenuation of the number of complexes was monitored as a function of laser frequency, yielding photofragmentation spectra.

Relative photo-absorption cross-sections. In the present experiment, a confined ensemble of some thousand C_{60}^+ -He complexes, $N(P)$, where P is the laser power measured at the exit of the instrument, is exposed to a continuous-wave diode laser. As can be seen from Fig. 3, the effective power density in the trap (diameter of the ion cloud is 0.8 cm) can be varied over a wide range. The observed attenuation curves have been fitted with the exponential function, $N(P) = N_0 \exp(-P/P_0)$. The characteristic power, P_0 , is a measure of the relative fragmentation cross-section of the complex. It is safe to assume that the relative fragmentation cross-section is equal to the relative absorption cross-section, σ_{rel} , and provides a reliable value for the absorption of C_{60}^+ itself. The results shown in Fig. 3 indicate that all complexes absorb at $9,632.7 \pm 0.1$ Å or $9,577.5 \pm 0.1$ Å. This supports the argument that, after many collisions with helium in the trap, only one structural isomer remains, as discussed in the main text. The relative cross-sections for four bands are presented in Table 1. Calculation of absolute cross-section values relies on the assumption that all ions uniformly explore the trapping volume. An attractive patch potential on one of the 22-pole electrodes, however, may cause the ions to stay outside the laser beam for longer than assumed. This leads to a large uncertainty in the absolute numbers and therefore only relative photo-absorption cross-sections are reported. The influence of power broadening is shown in Extended Data Fig. 1.



Extended Data Figure 1 | Influence of laser power on the 9,577.5 Å band. Gas-phase spectrum recorded by monitoring the depletion on the C_{60}^{+} -He mass channel using 1.5 mW (black) and 14 mW (red). Gaussian fits to

experimental data (circles) are represented by solid lines, and give FWHMs of $2.5 \pm 0.2 \text{ \AA}$ and $4.1 \pm 0.2 \text{ \AA}$ at 1.5 mW and 14 mW, respectively. The blue dashed line shows a Gaussian with a FWHM of 2.5 Å.

Quantum-dot-in-perovskite solids

Zhijun Ning^{1†*}, Xiwen Gong^{1*}, Riccardo Comin^{1*}, Grant Walters¹, Fengjia Fan¹, Oleksandr Voznyy¹, Emre Yassitepe¹, Andrei Buin¹, Sjoerd Hoogland¹ & Edward H. Sargent¹

Heteroepitaxy—atomically aligned growth of a crystalline film atop a different crystalline substrate—is the basis of electrically driven lasers, multijunction solar cells, and blue-light-emitting diodes^{1–5}. Crystalline coherence is preserved even when atomic identity is modulated, a fact that is the critical enabler of quantum wells, wires, and dots^{6–10}. The interfacial quality achieved as a result of heteroepitaxial growth allows new combinations of materials with complementary properties, which enables the design and realization of functionalities that are not available in the single-phase constituents. Here we show that organohalide perovskites and preformed colloidal quantum dots, combined in the solution phase, produce epitaxially aligned ‘dots-in-a-matrix’ crystals. Using transmission electron microscopy and electron diffraction, we reveal heterocrystals as large as about 60 nanometres and containing at least 20 mutually aligned dots that inherit the crystalline orientation of the perovskite matrix. The heterocrystals exhibit remarkable optoelectronic properties that are traceable to their atom-scale crystalline coherence: photoelectrons and holes generated in the larger-bandgap perovskites are transferred with 80% efficiency to become excitons in the quantum dot nanocrystals, which exploit the excellent photo-carrier diffusion of perovskites to produce bright-light emission from infrared-bandgap quantum-tuned materials. By combining the electrical transport properties of the perovskite matrix with the high radiative efficiency of the quantum dots, we engineer a new platform to advance solution-processed infrared optoelectronics.

Heteroepitaxy has so far largely relied on vacuum methods such as molecular-beam epitaxy, atomic-layer epitaxy, and metallo-organic vapour-phase epitaxy^{1,2}. These methods have been used to test and refine theoretical predictions of the conditions under which crystalline coherence is preserved even in the presence of a mismatch in the native lattices (strained-layer epitaxy). The result is a vast body of theory, knowledge, and practice regarding vapour-phase epitaxy. Since interfacial defects are rare at suitably designed hetero-interfaces, highly efficient luminescent materials have been created that have enabled efficient electrically injected lasers and light-emitting diodes for fibre-optic communications and high-efficiency lighting^{3–5}.

The past two decades have seen the rapid rise of soft condensed matter, often in the form of solution-processed semiconductors based on organic molecules, polymers, and colloidal nanoparticles (plates, wires, and dots)^{6–10}. At the same time, bulk organohalide semiconductor perovskites that exhibit large and perfect crystalline domains have improved in size, properties, and performance. These materials have enabled the development of the perovskite solar cell^{11–13}.

We investigated whether, under the right conditions, heteroepitaxial alignment could be produced between bulk perovskite semiconductors and quantum-tuned nanoparticles. Thin films consisting of multi-material, yet internally epitaxially aligned, heterocrystals—complex crystals we term poly-heterocrystalline (PHC) solids—could perhaps combine, without undue photocarrier loss, the desired properties of each phase: the excellent charge carrier transport properties of

the bulk perovskites, and the quantum-tuned infrared luminescence of the quantum-dot phase.

Solution processing is one approach to the application of organic perovskites. For example, methylammonium lead iodide perovskite (MAPbI₃) has led to the most efficient solution-processed solar cells reported so far. And, colloidal quantum dots (CQDs) based on PbS exhibit highly tunable photophysical properties, including in the near- and short-wavelength infrared, which the perovskites have so far failed to access. In terms of compatibility, MAPbI₃ and PbS possess related crystal structures (tetragonal and rock-salt, respectively), each having a six-coordinated Pb atom, and with Pb–Pb distances (MAPbI₃, 6.26 Å; PbS, 5.97 Å) that are within 4.6% of one another¹¹.

Consequently, we proposed that, if formed from a mixed solution phase into a solid phase (Fig. 1), then these two materials could potentially form PHC solids exhibiting optoelectronic properties that take advantage of the best features of each. Structurally, the PbS-CQD lattice matches well with the perovskite structure both three-dimensionally (Fig. 1a) and two-dimensionally (Fig. 1b). Studies have previously elucidated the synthesis and properties of composite organic/inorganic crystalline hybrids¹⁴; here we aim to engineer a heterocrystalline solid-state solution between two different, but structurally affine, materials. Specifically, we aim to combine the strong luminescent efficiency of the quantum dots with long-range carrier transport in the perovskite matrix.

To implement atomic-level coherence between the inorganic PbS and the organometallic MAPbI₃ phase, we require an organic-ligand-free strategy to cap the PbS nanoparticles. Colloids have recently been synthesized that were stabilized not using traditional aliphatic ligands, but rather using halide ligands that were introduced from the perovskite precursor methylammonium iodide^{15,16}. We use this method to prepare our CQDs.

Structural affinity is necessary but not sufficient to produce epitaxial bonding between the two phases; energetic considerations must also be taken into account. Here we use density functional theory (DFT) to study the interface-formation energy between the PbS (100) plane and the perovskite (110) plane (Extended Data Fig. 1)¹⁷. The interfacial energy is less than 10 meV Å^{–2}, suggesting that the growth of perovskite on PbS at room temperature is nearly as feasible as homoepitaxy of PbS on PbS or perovskite on perovskite. DFT further reveals that the epitaxial three-dimensional embedding of PbS CQDs inside a perovskite matrix is achieved without the formation of interfacial defects (Fig. 1c, d); that is, the bandgap is predicted to remain open, with no in-gap defects predicted (Extended Data Fig. 2).

We produced a PHC solid as follows. We first exchanged the organic ligands on the PbS CQDs for short halide (iodide in this case) anionic ligands (Extended Data Fig. 3a). We then mixed these inorganically terminated CQDs with PbI₂ dissolved in butylamine; this solvent strategy is crucial to maintaining passivation of the CQD surface following film growth. Since PbI₂ forms a complex with iodide and binds to the CQD surface, PbI₂ and CQDs were, as expected, able to mix in solution. (Transmittance spectra of the colloid showed no loss

¹Department of Electrical and Computer Engineering, University of Toronto, 35 St George Street, Toronto, Ontario M5S 1A4, Canada. †Present address: School of Physical Science and Technology, ShanghaiTech University, Haikou Road 100, 201210 Shanghai, China.

*These authors contributed equally to this work.

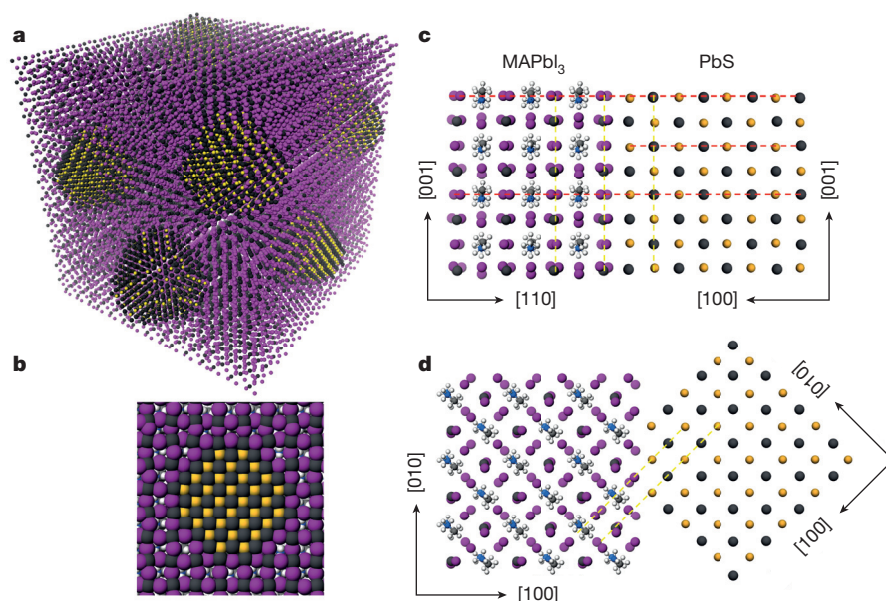


Figure 1 | Theoretical model of perovskite epitaxial growth on CQDs. **a**, Three-dimensional atomistic model of CQDs in a perovskite matrix. **b**, Cross-section (two-dimensional view) of a single CQD in perovskite. Here and in **c**, **d**, colours represent the following: grey, lead; purple, iodine; yellow, sulfur. All facets of CQDs match well with perovskite. **c**, **d**, Modelling of PbS and MAPbI₃ crystal structure and their interface, showing that perovskite matches well with PbS in both the X-Z plane (**c**) and the X-Y plane (**d**). The red dashes show the unit cell size; the yellow dashes are guides to the eye for matching planes.

of optical transmission for photon energies below the CQD bandgap, indicating no increase in scattering and thus no appreciable solution-phase aggregate formation.) By controlling the ratio of PbI₂ and CQDs, we tuned the nominal CQD concentration—hereafter expressed as the ratio of CQD to total volume—across the range 0.2–29% (see Extended Data Table 1a for a full list of values used). To fabricate the CQD–MAPbI₃ films (Extended Data Fig. 3b, c), we used a previously established sequential (two-step) method¹¹, whereby the PbI₂–CQD films are initially deposited via spin-coating onto a glass substrate, and are then soaked in a CH₃NH₃I/isopropanol solution. X-ray photoelectron spectroscopy (XPS) and Rutherford backscattering spectroscopy (RBS) were used to confirm the presence of both PbS and MAPbI₃

in the final film (Extended Data Fig. 4a–d). We detected no trace of amine ligands associated with the solvent (Extended Data Fig. 4e). RBS (Extended Data Fig. 4f) was used to estimate the CQD volume percentage as a function of nominal CQD:MAPbI₃ ratio.

High-resolution transmission electron microscopy (HRTEM) was used to characterize the crystal structure and orientation of both CQDs and perovskite matrix for a sample with a nominal CQD volume percentage of 4%. Since MAPbI₃ is an organic/inorganic hybrid material with lower density than PbS, it has much lower contrast than do inorganic CQDs, which facilitates differentiation between the perovskite and PbS phases. The higher contrast of CQDs was confirmed using high-angle annular dark-field scanning transmission electron

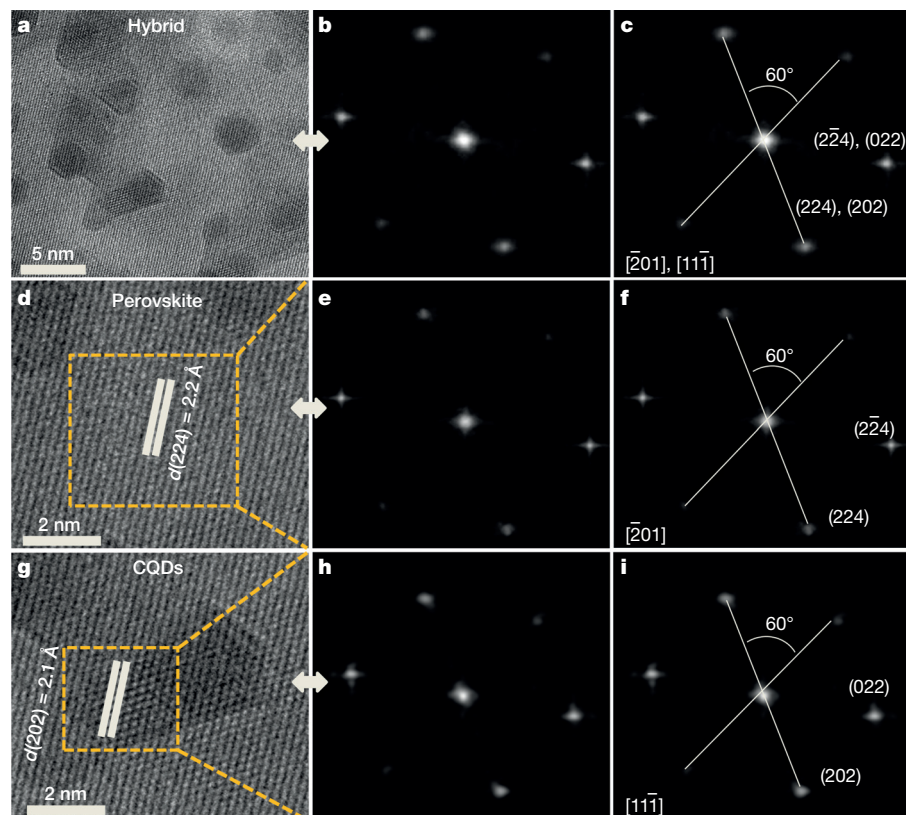


Figure 2 | HRTEM images and their FFTs. **a–c**, CQD–perovskite hybrid (CQD volume percentage is 3.9%); **d–f**, perovskite (MAPbI₃); **g–i**, CQDs. Images in **b**, **e**, and **h** are FFTs of the HRTEM images shown in **a**, and the yellow boxed regions in **d**, and **g**, respectively, and are the same as those in **c**, **f**, and **i**, but without the overlaid text. For perovskite, the lattice fringes are indexed to the (224) planes. For the CQDs, the lattice fringes are indexed to the (022) planes. In real space, the d -spacing value of the (224) plane of perovskite ($d(224) = 2.2 \text{ \AA}$) matches well with that of the (022) plane of PbS ($d(022) = 2.1 \text{ \AA}$). This is consistent with the modelling, which shows that the (100) plane of PbS matches well with the (110) plane of perovskite. The FFT images in **f** and **i** show the same intersection angles (60°) for perovskite (between the $\bar{2}24$ and (224) planes) and CQDs (between the (022) and (202) planes), indicating that they match not only two dimensionally, but also three dimensionally, which confirms that the perovskite and the CQDs have the same orientation. Adjustments of the brightness and contrast have been made to the FFT images for ease of visualization.

microscopy (HAADF-STEM), a direct probe of density variations in the material (Extended Data Fig. 5, Methods). The contrast between CQDs and perovskite is clearly observed in the large field-of-view HRTEM map of Fig. 2a; Fig. 2b shows the corresponding fast Fourier transform (FFT), with lattice planes indexed as indicated in Fig. 2c. The contributions from the perovskite matrix and the quantum dots are distinguished by analysing the subregions shown in Fig. 2d, g for the perovskite and quantum dots, respectively. Well defined lattice fringes with 2.2 Å separation (Fig. 2d) are indexed to the (224) plane of the matrix, as confirmed by the corresponding FFT images (Fig. 2e, f), which show that the real-space HRTEM image is projected along the $[201]$ zone axis. We also measured the lattice fringes of the CQDs, which show the same orientation and are indexed as the (022) plane of PbS (2.1 Å, Fig. 2g–i). This is consistent with simulations that show that the (100) facet of PbS matches well with the (110) facet of perovskite. As demonstrated by the FFT, the angles between $(2\bar{2}4)$ and $[201]$ of perovskite are the same as between (022) and $[11\bar{1}]$ of PbS. Hence, the orientation relationships between MAPbI₃ and CQDs are identified: $(022)_{\text{PbS}} \parallel (2\bar{2}4)_{\text{MAPbI}_3}$; $[11\bar{1}]_{\text{PbS}} \parallel [201]_{\text{MAPbI}_3}$. From this we conclude that the matrix and CQDs have the same orientation, which is observed in real-space microscope images and confirmed using a

FFT. This analysis demonstrates epitaxial alignment between perovskite and CQDs. Even a slight misalignment would be revealed by the experimental maps, as suggested by transmission electron microscopy (TEM) simulations of epitaxially oriented versus epitaxially misoriented PHCs (Extended Data Fig. 3d, e)¹⁸.

The optical-absorption spectra for pure CQD films and the new ‘dot-in-perovskite’ (hybrid) PHCs are shown in Fig. 3a. (Absorption and emission spectra for pure CQD and pure perovskite films are also reported in Extended Data Fig. 6a, b.) Absorption signatures corresponding to each constituent, perovskite and CQDs, are apparent in the hybrid films (Fig. 3a). The photoluminescence spectra are very similar to those of CQDs in film (Fig. 3b), indicating that the photophysical properties of CQDs are kept intact when they are incorporated into the perovskite matrix. As the CQD concentration increases, the absorption spectrum redshifts, consistent with increased inter-dot interaction (potentially partial dot fusion at these high concentrations), which reduces the extent of quantum confinement and thus shrinks the bandgap. This shrinking of the bandgap affects absorption and luminescence to the same extent; it does not, to leading order, impact the Stokes shift. A second effect of increased CQD concentration is increased long-range excitonic and carrier transport from dot to dot,

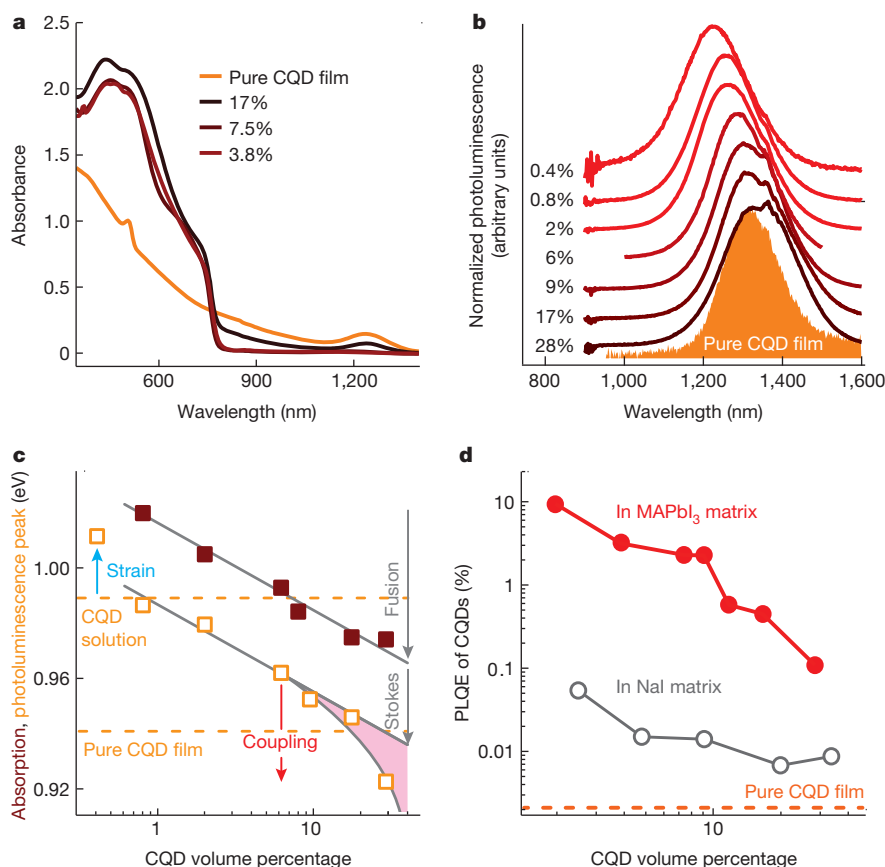


Figure 3 | Photophysical response of the CQD–perovskite hybrid.

a, b, Matrix-tuned absorption (**a**) and photoluminescence (**b**) of the CQD–perovskite hybrid. The different curves correspond to different CQD volume percentages, as labelled. After the growth of perovskite, a representative absorption edge around 770 nm appears, whereas both absorption and photoluminescence CQD exciton peaks are preserved, indicating that the properties of CQDs remain unaltered. **c,** The absorption (red filled squares) and photoluminescence exciton peak positions (orange open squares) for the CQDs as a function of CQD:MAPbI₃ ratio (illustrated via the CQD volume percentage). Orange dashed lines mark the photoluminescence emission energy for pure CQD films and solutions. Different mechanisms governing the concentration-dependent absorption and emission are also highlighted: CQD fusion causes a redshift, which increases as CQD loading increases (grey arrow ‘Fusion’); inter-dot coupling causes a redshift at high CQD

concentrations (pink arrow ‘Coupling’ and shading); and CQD lattice strain that is driven by epitaxial effects, which only has an effect for low CQD concentrations, causes a blueshift (blue arrow ‘Strain’). Grey lines are guides to the eye. The concentration-independent Stokes shift causes a redshift (grey arrow ‘Stokes’). **d,** PLQE of CQDs in matrix. For lower CQD concentration the PLQE is increased, indicating that CQDs are well dispersed in the matrix and carrier dissociation between CQDs is blocked. In the MAPbI₃ matrix, the PLQE is two orders of magnitude higher than that in a NaI matrix, which reveals that a lattice-mismatched matrix (8.8% mismatch for NaI) cannot grow on PbS. This finding confirms that epitaxial growth of matrix on CQDs is critical for CQD surface passivation. The PLQE of a solution-exchanged CQD film (orange dashed line) is shown for comparison. The PLQE of CQDs in the MAPbI₃ matrix is approximately 3,000 times greater than in the CQD film.

which is present only at the highest concentrations of CQDs. Through this process, excitons and carriers are transferred to the smallest-gap dots in a population whose lowest-lying quantum-confined levels are inhomogeneously broadened. The increased long-range excitonic and carrier transport does not affect absorption, but redshifts the luminescence, and thus increases the Stokes shift^{19,20}. In addition to the mechanisms of partial fusion and increased long-range transport that contribute redshifts, a blueshift (even relative to dots in solution) may also be contributed especially at low CQD loading: tensile strain caused by the lattice mismatch between CQDs and the perovskite slightly expands the PbS lattice, leading to an increase in bandgap²¹ (Fig. 3c). This tensile strain acts to the same extent on absorption and emission; see also Extended Data Fig. 7a, b and Methods.

We investigated the CQD photoluminescence quantum efficiency (PLQE)—the ratio of the number of emitted and absorbed photons—for a series of samples with different CQD concentrations (expressed as volume percentages). Initially we used an excitation wavelength of 815 nm, which is below the absorption edge of the perovskite (Fig. 3d), to excite the quantum-dot phase directly and exclusively. The PLQE is highest in low-dot-concentration samples, consistent with minimal quenching associated with the dissociation of excitons into free carriers in the neighbouring quantum dots²². The PLQE reaches 9.3%, which is two orders of magnitude higher than that in the pure CQD film. As a control to study the importance of lattice matching on photophysical properties, we changed the matrix to sodium iodide (NaI, rock-salt structure with an 8.8% lattice mismatch with PbS): the PLQE of CQDs in a NaI matrix is almost two orders of magnitude lower than it is in a perovskite matrix (Fig. 3d).

To elucidate the physical origins of enhanced PLQE for dots embedded in a perovskite matrix, we used DFT to investigate whether CQD surfaces can in principle be passivated using the perovskite matrix. Both the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) are located in the core of PbS CQDs, and the bandgap is trap free. In contrast, for a CQD that is misaligned

with the perovskite matrix, trap states are induced at the CQD surface, mostly on the conduction-band side (Extended Data Fig. 2c, d, f).

In addition to providing passivation, the matrix–dot interface is the critical enabler of any charge carrier injection into the CQDs. To study carrier transfer across this interface, we relied on CQDs that are approximately 4 nm in diameter (bandgap of 1 eV) whose LUMO energy level is predicted to lie below the conduction band of the perovskite and whose HOMO energy level should reside above the valence-band edge of the perovskite (Fig. 4a).

By generating excitons in the CQDs only (via photoexcitation using photon energies lower than the perovskite bandgap), and in separate studies where we photogenerate principally in the perovskite matrix, we investigate carrier transfer to, and across, the CQD–perovskite interface. Photoluminescence excitation spectra (Fig. 4b) reveal enhanced CQD emission when the perovskite is excited ($\lambda_{\text{excitation}} < 780 \text{ nm}$)²³. Increased perovskite content drives a higher CQD photoluminescence intensity when the excitation occurs in the perovskite absorption region. This result qualitatively indicates efficient charge diffusion to, and across, the perovskite–CQD interface. Carrier transfer also occurs in other composite nanomaterials^{22,24}, such as those made up of small-bandgap–CQD inclusions in a large-bandgap–CQD matrix (‘dots-in-dots’ films); however, we found that the use of a perovskite matrix provided superior passivation of the quantum dot centres, as evidenced by the nearly two orders of magnitude higher PLQE compared to the same dots in a NaI matrix, and by the fivefold increase in carrier lifetime compared to ‘dots-in-dots’ films (Extended Data Fig. 8a, b).

The signatures of carrier transfer from the matrix to the quantum dots are also evident in the photoluminescence emission from MAPbI₃ in PHC films, as a function of CQD volume percentage (Fig. 4c). In particular, for volume percentages greater than 0.4%, we observe a complete quenching of the perovskite photoluminescence signal, consistent with a highly-efficient removal of photogenerated carriers from the matrix and transfer into the dots. At higher CQD concentrations—in the regime of low densities of injected carriers per dot, where

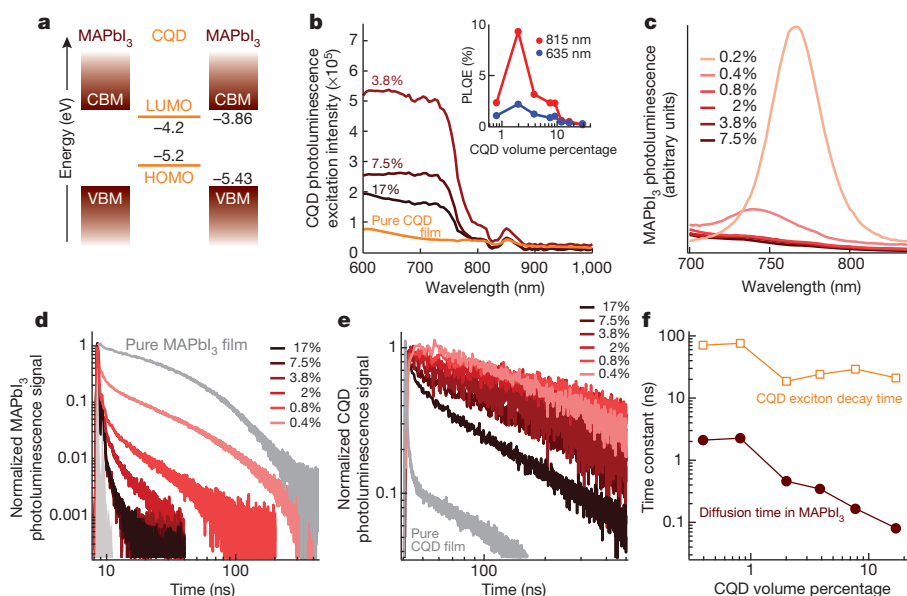


Figure 4 | Carriers transfer from perovskite to CQDs. **a**, Electronic band structure of CQDs and perovskite. The LUMO level of CQD is lower than the conduction band minimum (CBM) of perovskite and the HOMO level of CQD is higher than the valence band maximum (VBM) of perovskite, making the transfer of carriers from perovskite to CQDs energetically favoured. **b**, Photoluminescence excitation spectra of CQDs. In the perovskite absorption region ($\lambda < 780 \text{ nm}$), the photoluminescence excitation intensity increases as perovskite content increases, indicating that carriers excited in perovskite are transferred to CQDs. However, considering the partial carrier loss (via recombination) in perovskite, the PLQE of the hybrid is lower than the

internal PLQE of CQD when long excitation wavelength is used (inset). **c**, Photoluminescence spectra of MAPbI₃ for PHC films. **d–f**, Transient photoluminescence emission from MAPbI₃ (**d**) and CQDs (**e**), and corresponding time constants (**f**). As the CQD:MAPbI₃ ratio is reduced (the CQD volume percentage decreases), the initial photoluminescence decay time of both perovskite and CQDs becomes longer, as a result of longer transit time in the matrix and reduced carrier dissociation across the CQDs, respectively. The different curves in **b–e** correspond to different CQD volume percentages, as labelled.

radiative recombination is dominant over higher-order (for example, Auger) processes—we quantitatively estimate the carrier transfer efficiency from perovskite to CQDs on the basis of photoluminescence and absorption data (Extended Data Table 1b, Methods). The transfer efficiency attains a value exceeding 80% at the highest CQD loading (28% CQD volume percentage).

We investigated the carrier dynamics in the PHC using photoluminescence decay (Fig. 4d, e). We used a two-component decay model to analyse transient photoluminescence. The fast decay that is transient in CQDs is attributed to fast exciton dissociation into neighbouring CQDs, whereas the longer dynamics are ascribed to radiative recombination. The comparison between the PHC and the pure CQD photoluminescence transients in Fig. 4e shows a rapid slowdown of exciton-dissociation dynamics as soon as the CQDs are embedded in the perovskite matrix (from less than 1 ns in pure CQD to 20 ns in PHC at high CQD loading). Furthermore, the fast transient component increases from 20 ns to 70 ns as the CQD concentration is reduced (Fig. 4f) and, correspondingly, the inter-dot separation is increased, consistent with the expected suppression in the exciton-dissociation rate for more isolated dots. The perovskite carrier dynamics (Fig. 4d) show a rapid quenching of the photoluminescence emission, from 25 ns in pure MAPbI₃ to about 2 ns in the PHC film at the lowest CQD loading. The photoluminescence lifetime associated with emission from the perovskite phase in ‘quantum-dot-in-perovskite’ solids is highest (2 ns) at the lowest CQD concentration (Fig. 4f). At this lowest concentration, the inter-dot spacing is an estimated 40 nm, and this relatively larger separation increases the time that is required for photocarriers to diffuse from their point of photogeneration in the perovskite phase to the point of capture into a quantum dot (Extended Data Fig. 9). We use the term ‘charge carrier’ when referring to photo-excitations in the perovskite phase because prior studies have revealed that photogenerated excitons in perovskite separate to become free carriers within about 2 ps after generation²⁵. This analysis does not preclude the possibility that some fraction of excitations could transfer as excitons from perovskite to dot, especially for those excitons that are generated very close to a quantum dot.

This work reports an *in situ* epitaxial growth process of perovskite on quantum-dot surfaces. The quantum dots are effectively passivated by the matrix, without the need for conventional organic ligands, provided lattice-matching conditions are satisfied. The perovskite matrix also provides excellent carrier transport for the CQDs, owing to its superior diffusion length. The present approach paves the way for new strategies to enhance the performance of CQD-based optoelectronic devices, and to extend the spectral diversity of perovskite-based materials towards the infrared.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 December 2014; accepted 7 May 2015.

1. Pohl, U. W. *Epitaxy of Semiconductors: Introduction to Physical Principles* (Springer, 2013).
2. McCray, W. P. MBE deserves a place in the history books. *Nature Nanotechnol.* **2**, 259–261 (2007).
3. Huang, M. H. *et al.* Room-temperature ultraviolet nanowire nanolasers. *Science* **292**, 1897–1899 (2001).

4. Choi, K., Kako, S., Holmes, M. J., Arita, M. & Arakawa, Y. Strong exciton confinement in site-controlled GaN quantum dots embedded in nanowires. *Appl. Phys. Lett.* **103**, 171907 (2013).
5. Adhikari, H., Marshall, A. F., Chidsey, C. E. D. & McIntyre, P. C. Germanium nanowire epitaxy: shape and orientation control. *Nano Lett.* **6**, 318–323 (2006).
6. Diao, Y. *et al.* Solution coating of large-area organic semiconductor thin films with aligned single-crystalline domains. *Nature Mater.* **12**, 665–671 (2013).
7. Shirasaki, Y., Supran, G. J., Bawendi, M. G. & Bulović, V. Emergence of colloidal quantum-dot light-emitting technologies. *Nature Photon.* **7**, 13–23 (2013).
8. Graetzel, M., Janssen, R. A. J., Mitzi, D. B. & Sargent, E. H. Materials interface engineering for solution-processed photovoltaics. *Nature* **488**, 304–312 (2012).
9. Talapin, D. V., Lee, J., Kovalenko, M. V. & Shevchenko, E. V. Prospects of colloidal nanocrystals for electronic and optoelectronic applications. *Chem. Rev.* **110**, 389–458 (2010).
10. Ma, W., Luther, J. M., Zheng, H., Wu, Y. & Alivisatos, A. P. Photovoltaic devices employing ternary PbS_{1-x}Se_x nanocrystals. *Nano Lett.* **9**, 1699–1703 (2009).
11. Burschka, J. *et al.* Sequential deposition as a route to high-performance perovskite-sensitized solar cells. *Nature* **499**, 316–319 (2013).
12. Tan, Z. *et al.* Bright light-emitting diodes based on organometal halide perovskite. *Nature Nanotechnol.* **9**, 687–692 (2014).
13. Stranks, S. D. *et al.* Electron-hole diffusion lengths exceeding 1 micrometer in an organometal trihalide perovskite absorber. *Science* **342**, 341–344 (2013).
14. Mitzi, D. B., Chondroudis, K. & Kagan, C. R. Organic-inorganic electronics. *IBM J. Res. Dev.* **45**, 29–45 (2001).
15. Ning, Z., Dong, H., Zhang, Q., Voznyy, O. & Sargent, E. H. Solar cells based on inks of n-type colloidal quantum dots. *ACS Nano* **8**, 10321–10327 (2014).
16. Dirin, D. N. *et al.* Lead halide perovskites and other metal halide complexes as inorganic capping ligands for colloidal nanocrystals. *J. Am. Chem. Soc.* **136**, 6550–6553 (2014).
17. Leitsmann, R. *et al.* Adsorption mechanisms of fluorocarbon polymers at ultra low-*k* surfaces. *Surf. Sci.* **604**, 1808–1812 (2010).
18. Koch, C. T. *Determination of Core Structure Periodicity and Point Defect Density along Dislocations*. PhD thesis, Arizona State Univ. (2002).
19. Moroz, P. *et al.* Suppressed carrier scattering in CdS-encapsulated PbS nanocrystal films. *ACS Nano* **7**, 6964–6977 (2013).
20. Gao, Y. *et al.* Enhanced hot-carrier cooling and ultrafast spectral diffusion in strongly coupled PbSe quantum-dot solids. *Nano Lett.* **11**, 5471–5476 (2011).
21. Wise, F. W. Lead salt quantum dots: the limit of strong quantum confinement. *Acc. Chem. Res.* **33**, 773–780 (2000).
22. Choi, J. J. *et al.* Photogenerated exciton dissociation in highly coupled lead salt nanocrystal assemblies. *Nano Lett.* **10**, 1805–1811 (2010).
23. Kagan, C. R., Murray, C. B., Nirmal, M. & Bawendi, M. Electronic energy transfer in CdSe quantum dot solids. *Phys. Rev. Lett.* **76**, 1517–1520 (1996).
24. Xu, F. *et al.* Efficient exciton funneling in cascaded PbS quantum dot superstructures. *ACS Nano* **5**, 9950–9957 (2011).
25. Ponceca, C. S. *et al.* Organometal halide perovskite solar cell materials rationalized: ultrafast charge generation, high and microsecond-long balanced mobilities, and slow recombination. *J. Am. Chem. Soc.* **136**, 5189–5192 (2014).

Acknowledgements This publication is based, in part, on work supported by an award (KUS-11-009-21) from the King Abdullah University of Science and Technology (KAUST), by the Ontario Research Fund Research Excellence Program and by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Computations were performed using the BlueGene/Q supercomputer at the SciNet HPC Consortium provided through the Southern Ontario Smart Computing Innovation Platform (SOSCIIP). E.Y. acknowledges support from an FAPESP-BEPE (14/18327-9) fellowship. The authors thank L. Levina for assistance in CQD synthesis, E. Beauregard for assistance in PHC synthesis, Z. Yang and M. Adachi for discussions, and E. Palmiano, R. Wolowiec and D. Kopilovic for their help during the course of study.

Author Contributions Z.N., X.G., R.C. and E.H.S. designed and directed the study. Z.N. and X.G. contributed to all the experimental work. R.C. carried out the photoluminescence lifetime and excitation measurements and analysis. G.W. and S.H. performed the PLQE and the carrier transfer efficiency study. F.F. and E.Y. did the TEM measurement and FFT analysis. O.V. carried out the TEM simulation and XPS measurement. O.V. and A.B. performed the DFT simulation. Z.N., X.G., R.C., and E.H.S. wrote the manuscript. All authors commented on the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.H.S. (ted.sargent@utoronto.ca).

METHODS

CQD synthesis and solution ligands exchange. Colloidal quantum dots (CQDs) were synthesized using methods previously reported²⁶. For iodide ligand exchange, 3 ml of CQDs dispersed in octane (10 mg ml^{-1}) were added into 3 mL of dimethylformamide (DMF) solution containing 350 mg of PbI_2 and 150 mg of $\text{CH}_3\text{NH}_3\text{I}$. After stirring for 10 min, CQDs had transferred from the top octane phase to the bottom DMF. After removing the octane, we washed the CQD solution three more times using octane to remove the organic residue. Subsequently, the CQDs were precipitated by the addition of toluene. The nanoparticles were dispersed in butylamine for film fabrication.

Poly-heterocrystalline (PHC) solids fabrication. A given amount of PbI_2 (with one quarter weight ratio of $\text{CH}_3\text{NH}_3\text{I}$) was added into the solution for perovskite growth. A spin-coating process was used for film fabrication (spin speed of 6,000 r.p.m.). The film was then annealed at 70°C for 10 min in a N_2 glovebox. For perovskite growth, methylammonium iodide solution (10 mg ml^{-1} in isopropanol) was dropped and left on the film for 30 s. After removing the solution by spin coating (spin speed 6,000 r.p.m.), the film was soaked in pure isopropanol for 10 s, and then a spin-coating process was performed to remove the residual solvent. Finally, the film was annealed again at 70°C for 10 min in a N_2 glovebox.

Density functional theory (DFT) simulations. Calculations were performed within the DFT formalism using a Perdew–Burke–Ernzerhof²⁷ generalized gradient approximation exchange correlation functional. All calculations were performed using the CP2K²⁸ package with mixed Gaussian and plane-wave basis set, using the molecularly optimized MOLOPT²⁸ double ζ -valence polarized (mDZVP) basis set implemented in CP2K code, which has very small basis set superposition errors in gas and condensed phases^{17,29–31}. The charge density cut-off was 300 Ry, which is suitable for the Goedecker–Teter–Hutter pseudopotentials³². Spin polarized (local-spin-density approximation) and spin-unpolarized calculations (local-density approximation) were performed in the cases of odd and even numbers of electrons, respectively. The structural minimization was performed using a Broyden–Fletcher–Goldfarb–Shanno algorithm³³. Surface slabs were modelled as (110) MAPbI_3 of tetragonal structure and (100) cubic PbS with seven monolayers each. 100 \AA of vacuum was added on top of the slab. Dipole correction was used to remove artificial dipole–dipole interaction across periodic images as implemented in the CP2K v2.5. A 3×3 ($26.88 \text{ \AA} \times 26.88 \text{ \AA}$) supercell was used in the X–Y plane.

The interfacial energy was computed using the relation³⁴

$$E_{\text{inter}}^{\text{MAPbI}_3-\text{PbS}} = E_s(\text{MAPbI}_3) + E_s(\text{PbS}) - E_{\text{ad}}(\text{MAPbI}_3-\text{PbS})$$

where E_s represents the surface energies of the corresponding surfaces (MAPbI_3 and PbS), and $E_{\text{ad}}(\text{MAPbI}_3-\text{PbS})$ is the adsorption energy of the MAPbI_3 on PbS . The interfacial energy describes the energy required to cut the PbS and MAPbI_3 bulk and combine the obtained free surfaces to form the MAPbI_3 – PbS interface.

The estimated surface energy of $\text{PbS}(100)$ is 9 meV \AA^{-2} , which is in close agreement with previously reported³⁵ values of 12 meV \AA^{-2} . The MAPbI_3 surface energy is^{36,37} 9 meV \AA^{-2} . The adsorption energy is calculated via

$$E_{\text{ad}}(\text{MAPbI}_3-\text{PbS}) = E_{\text{slab}}(\text{MAPbI}_3) + E_{\text{slab}}(\text{PbS}) - E_{\text{slab}}(\text{MAPbI}_3-\text{PbS})$$

where E_{slab} is the slab energy of the corresponding bare MAPbI_3 or PbS slab, and $E_{\text{slab}}(\text{MAPbI}_3-\text{PbS})$ is the energy of the combined PbS – MAPbI_3 slab. The adsorption energy is calculated to be about 12 meV \AA^{-2} , which gives a value of the interfacial energy of the order of the 6 meV \AA^{-2} . The value obtained for the interfacial energy, which is smaller than that of free surfaces, is indicative of the matched interface being favoured over broken bonds at the interface.

Calculations to investigate a PbS CQD embedded in a perovskite matrix were performed using a 50.6 \AA^3 box that contained about 3,000 atoms. The relaxed lattice constant of the perovskite was imposed on the unit cell, corresponding to approximately 5% tensile strain on PbS . A PbS CQD of about 3 nm in diameter fully relaxed in vacuum was placed inside a manually prepared cavity in perovskite providing an epitaxial matching. Fully quantum mechanical total-energy minimization was performed on the combined system.

We find that the PbS lattice constant is changed by only 0.5% strain, instead of by the imposed 5% difference between its lattice constant and that of the perovskite (Extended Data Fig. 7). The preponderance of the lattice constant change is accommodated in the perovskite in a thin layer within one or two lattice constants near the interface. This can be explained by the fact that PbS is much stiffer than perovskite: the bulk modulus of PbS is $B_{\text{PbS}} = 53 \text{ GPa}$ (ref. 38), whereas for perovskite it is $B_{\text{perovskite}} = 12.2 \text{ GPa}$ (ref. 39). Calculated bandgaps for a strained PbS 3-nm CQD in vacuum are: $E_g[\text{no strain}] = 1.06 \text{ eV}$, and $E_g[0.5\% \text{ tensile strain}] = 1.08 \text{ eV}$, consistent with deformation potentials derived from temperature dependent bandgaps for these CQD sizes²¹.

This bandgap change corresponds to an approximately 24 nm blueshift, consistent with the excess blueshift observed experimentally in low-loading samples that is greater than the shift associated with low dot-to-dot communication.

Decoupling of perovskite and CQD contribution in TEM. We used high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) to image the dots in perovskite. The corresponding signal, which originates from electrons that are scattered incoherently at large angles, provides contrast between the heavier (denser) CQDs and the lighter MAPbI_3 matrix. Under these conditions, contributions from diffraction and phase are suppressed, and the contrast therefore depends mainly on the atomic number Z . Consequently, the denser PbS CQDs will scatter electrons more strongly, yielding a higher intensity at the detector than when imaging the perovskite crystal matrix. This contrast is seen in the HAADF-STEM map of a CQD–perovskite thin film with 3.9% CQD loading (Extended Data Fig. 5c), with brighter regions marking the presence of CQDs with diameters of about 5 nm. The absence of these intensity variations in a pure perovskite matrix is shown in Extended Data Fig. 5a, which shows a single perovskite grain (white region) in which no brightness fluctuations are observed. The corresponding HRTEM maps are shown in Extended Data Fig. 5b (pure perovskite film) and the inset of Extended Data Fig. 5c (CQD–perovskite thin film). In the latter, the HRTEM is only acquired in the region highlighted by the yellow circle, which is expanded in Extended Data Fig. 5d to provide a comparison with the HRTEM map shown in Fig. 2g. Notably, the CQDs appear as dark in HRTEM, demonstrating the validity of the framework used for the analysis discussed in the main text.

Extraction of transport efficiency from PLQE and photoluminescence decay. We define the total transfer efficiency (η_{tot}) and diffusion transport efficiency (η_{diff}) as

$$\eta_{\text{tot}} = \frac{n_{\text{trans}}}{n_{\text{all}}} \quad \text{and} \quad \eta_{\text{diff}} = \frac{n_{\text{diff}}}{n_{\text{all}}}$$

where n_{trans} is the number of charge carriers that are transferred into the CQDs from the perovskite, n_{diff} is the number of carriers that reach the CQDs by diffusing across the perovskite matrix, and n_{all} is the total number of photogenerated charge carriers in the perovskite.

Experimentally, η_{tot} is extracted from photoluminescence quantum efficiency (PLQE) measurement using a modified version of a method reported in the literature⁴⁰. We measured the photoluminescence from CQDs in matrix using two excitation wavelengths: a short wavelength that excites both CQDs and perovskite, and a long wavelength that only excites CQDs. The photoluminescence (PL) of CQDs in these two scenarios are

$$\text{PL}_{\text{CQDs,short}} = (A_{\text{CQDs,short}} + \eta_{\text{tot}} A_{\text{p,short}}) \times \text{PLQE}_{\text{CQDs}} \times I_{\text{ex,short}} \quad (1)$$

$$\text{PL}_{\text{CQDs,long}} = A_{\text{CQDs,long}} \times \text{PLQE}_{\text{CQDs}} \times I_{\text{ex,long}} \quad (2)$$

PL_{CQDs} and I_{ex} represent the photoluminescence yield from the CQDs (in photons per second) and the photon intensity of the excitation source (in photons per second), respectively. A_{CQDs} and A_{p} are the absorption of CQDs and perovskite, respectively. A_{CQDs} was determined by measuring the change in absorbance at the dot's excitonic peak and scaling the total material absorbance at the chosen wavelength by this factor. A_{p} was determined using the additive property of absorbances for a mixture. The Beer–Lambert law was used to convert the experimental absorption measurements to absorbances for use in these calculations. Owing to very low optical density for the absorption of CQDs in films with low CQD concentration, the values of A_{CQDs} were estimated using a linear scaling (according to the CQD concentration), on the basis of the measured A_{CQDs} for a film with 28% CQD concentration (highest loading). The PLQE is calculated using an integrating sphere and following a method described elsewhere⁴¹. The intensity of the excitation source was measured using a power meter (LaserStar). The wavelength-dependent responses of the visible and near-infrared photon detectors (for photoluminescence and PLQE measurements) were corrected using an Ocean Optics LS-1 calibration lamp with a known spectral profile. The ‘short’ and ‘long’ in the subscripts indicate the wavelength regime. From equations (1) and (2), we determine

$$\eta_{\text{tot}} = \left[\left(\frac{\text{PL}_{\text{CQDs,short}} \times I_{\text{ex,long}}}{\text{PL}_{\text{CQDs,long}} \times I_{\text{ex,short}}} \right) A_{\text{QDs,long}} - A_{\text{QDs,short}} \right] \frac{1}{A_{\text{p,short}}} \quad (3)$$

The CQD photoluminescence scans that are used to extract the total integrated (with respect to wavelength) photoluminescence intensity (normalized with the incident light intensity I_{ex}), $\text{PL}_{\text{CQDs}}/I_{\text{ex}}$, are plotted against wavelength in Extended Data Fig. 6c, d. Owing to the presence of a detector cut-off at long wavelength (about 1,380 nm), the integrated area of the left half-peak (shaded area in Extended Data

Fig. 6c, d) is evaluated as $\text{PL}_{\text{CQDs}} = \int_{1,000 \text{ nm}}^{\lambda_{\text{peak}}} \text{PL}(\lambda) d\lambda$ and then doubled to obtain the total photoluminescence. The absorption scans used to extract the absorption

coefficients A_{CQDs} and A_p are plotted in Extended Data Fig. 6e; the results are presented in Extended Data Table 1b.

Dynamic model for time-resolved photoluminescence. For the perovskite matrix, by considering bimolecular recombination as the dominant radiative relaxation pathway⁴², the photoluminescence intensity of perovskite is

$$I_{\text{PL},p}(t) = k_{\text{rad},p} [n_p(t)]^2 \quad (4)$$

where $k_{\text{rad},p}$ is the bimolecular recombination rate of perovskite, and $n_p(t)$ is charge carrier density in perovskite, whose time evolution is modelled as arising from two contributions: a fast exponential decay (carriers rapidly transferred to dots) plus a slower radiative decay

$$n_p(t) = a_1 e^{-k_{\text{ct}}t} + \frac{1}{k_{\text{rad},p}t + a_2} \quad (5)$$

where a_1 and a_2 are constants, and k_{ct} is the charge carrier transfer rate, which corresponds to the reciprocal of the average time for a carrier to reach (and get captured by) a CQD via diffusion in the matrix. From equations (4) and (5), the photoluminescence intensity of perovskite is

$$I_{\text{PL},p}(t) = k_{\text{rad},p} \left[a_1 e^{-k_{\text{ct}}t} + \frac{1}{k_{\text{rad},p}t + a_2} \right]^2 \quad (6)$$

For CQDs in matrix, both exciton recombination (triggered by the direct excitation of excitons in dots) and bimolecular recombination (due to free carriers injected from the matrix) contribute to photoluminescence:

$$I_{\text{PL},\text{CQDs}}(t) = k_{\text{ex-rad}} n_{\text{ex}}(t) + k_{\text{rad,CQDs}} [n_{\text{CQDs}}(t)]^2 \quad (7)$$

Here, $k_{\text{ex-rad}}$ is the exciton radiative recombination rate, $k_{\text{rad,CQDs}}$ is the bimolecular recombination rate of free carriers in CQDs, and n_{ex} and $n_{\text{CQDs}}(t)$ represent the exciton and injected free carrier density in CQDs, respectively. The exciton density and free carrier density are

$$n_{\text{ex}}(t) = n_{\text{ex}0} e^{-k_{\text{ex}}t} \quad (8)$$

$$n_{\text{CQDs}}(t) = \frac{1}{k_1 t + a_3} (1 - e^{-k_{\text{ct}}t}) \quad (9)$$

respectively, where $n_{\text{ex}0}$ represents the initial exciton density, k_{ex} and k_1 are the total recombination rate of excitons and free charge carrier constants, respectively, and a_3 is a constant.

Considering equations (7)–(9), the photoluminescence intensity of CQDs is

$$I_{\text{PL},\text{CQDs}}(t) = k_{\text{ex}} n_{\text{ex}0} e^{-k_{\text{ex}}t} + k_{\text{rad,CQDs}} \left[\frac{1}{k_1 t + a_3} (1 - e^{-k_{\text{ct}}t}) \right]^2 \quad (10)$$

The models given in equations (6) and (10) are used to fit the photoluminescence decay traces of CQDs and perovskite and extract the parameters that are most relevant for charge transport, loss, and recombination across the film.

Extended Data Fig. 10 shows the integrated (with respect to time) photoluminescence from the exciton term in equation (10)

$$\widetilde{\text{PL}}_{\text{ex}} = \int k_{\text{ex}} n_{\text{ex}0} e^{-k_{\text{ex}}t} dt$$

as well as the integrated (with respect to time) photoluminescence from free carrier recombination term in equation (10)

$$\widetilde{\text{PL}}_{\text{ca}} = \int k_{\text{rad,CQDs}} \left[\frac{1}{k_1 t + a_3} (1 - e^{-k_{\text{ct}}t}) \right]^2 dt$$

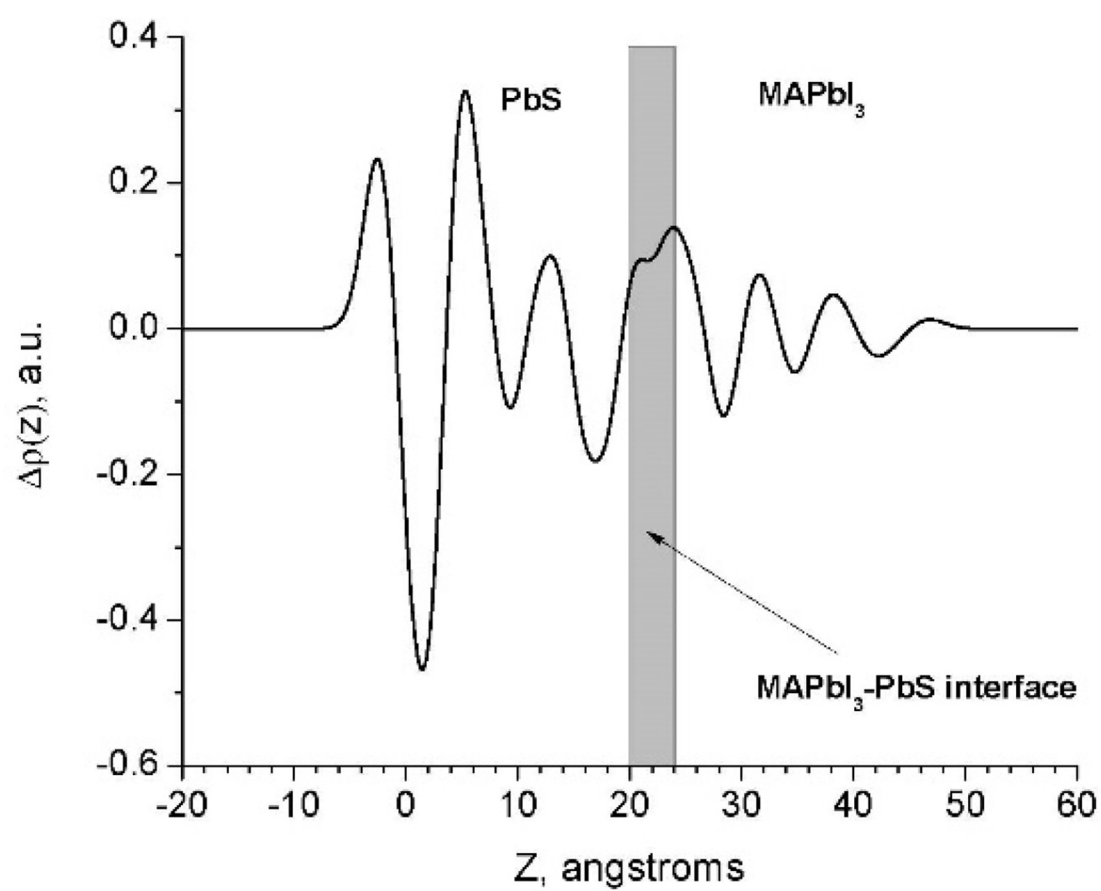
Estimate of carrier diffusion length in perovskite matrix. In general, the diffusion length L_D of carriers in any material may be calculated from the diffusion coefficient D and the carrier lifetime τ using

$$L_D = \sqrt{D\tau} \quad (11)$$

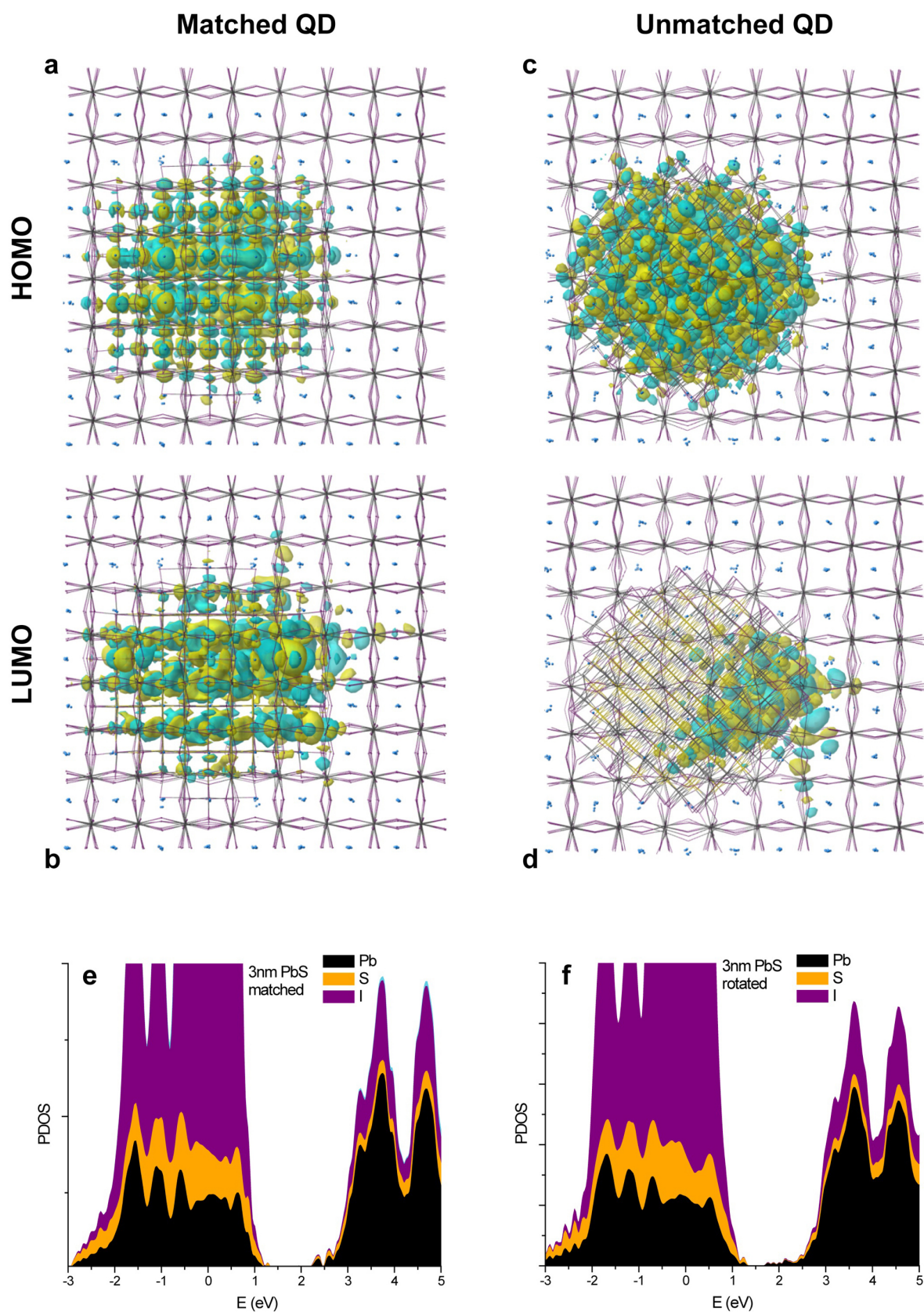
Using the experimentally determined carrier lifetime in the perovskite matrix (evaluated from the decay time of the perovskite photoluminescence emission; see Fig. 4d), we estimate the diffusion length for carriers in our PHC films for possible values of the diffusion coefficient (which cannot be directly measured) in the range previously determined in pure lead iodide perovskite films¹³. The corresponding data points for L_D as a function of CQD volume concentration are shown in Extended Data Fig. 9. This figure shows that the estimated values for L_D are comparable to the calculated average inter-dot spacing (red line), which suggests that the diffusion length in the matrix is ultimately determined by the average free space available for carriers to diffuse in the matrix before being transferred to the CQDs.

Code availability. CP2K is freely available from <http://www.cp2k.org/>.

26. Hines, M. A. & Scholes, G. D. Colloidal PbS nanocrystals with size-tunable near-infrared emission: observation of post-synthesis self-narrowing of the particle size distribution. *Adv. Mater.* **15**, 1844–1849 (2003).
27. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
28. VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **127**, 114105 (2007).
29. Takaluoma, T. T., Laasonen, K. & Laitinen, R. S. Molecular dynamics simulation of the solid-state topochemical polymerization of S_2N_2 . *Inorg. Chem.* **52**, 4648–4657 (2013).
30. Bork, N., Loukonen, V. & Vehkamäki, H. Reactions and reaction rate of atmospheric SO_2 and $\text{O}^-(\text{H}_2\text{O})_n$ collisions via molecular dynamics simulations. *J. Phys. Chem. A* **117**, 3143–3148 (2013).
31. Smecca, E., Motta, A., Fragal, M. E., Aleeva, Y. & Condorelli, G. G. Spectroscopic and theoretical study of the grafting modes of phosphonic acids on ZnO nanorods. *J. Phys. Chem. C* **117**, 5364–5372 (2013).
32. Hartwigsen, C., Goedecker, S. & Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **58**, 3641–3662 (1998).
33. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing* 3rd edn, Ch. 10.9 (Cambridge Univ. Press, 2007).
34. Liu, W., Li, J. C., Zheng, W. T. & Jiang, Q. Ni Al (110)/Cr (110) interface: a density functional theory study. *Phys. Rev. B* **73**, 205421 (2006).
35. Zhrebetskyy, D. et al. Hydroxylation of the surface of PbS nanocrystals passivated with oleic acid. *Science* **344**, 1380–1384 (2014).
36. Buin, A. et al. Materials processing routes to trap-free halide perovskites. *Nano Lett.* **14**, 6281–6286 (2014).
37. Haruyama, J., Sodeyama, K., Han, L. & Tateyama, Y. Termination dependence of tetragonal $\text{CH}_3\text{NH}_3\text{PbI}_3$ surfaces for perovskite solar cells. *J. Phys. Chem. Lett.* **5**, 2903–2909 (2014).
38. Bian, K., Bassett, W., Wang, Z. & Hanrath, T. The strongest particle: size-dependent elastic strength and Debye temperature of PbS nanocrystals. *J. Phys. Chem. Lett.* **5**, 3688–3693 (2014).
39. Feng, J. Mechanical properties of hybrid organic-inorganic $\text{CH}_3\text{NH}_3\text{BX}_3$ (B = Sn, Pb; X = Br, I) perovskites for solar cell absorbers. *APL Mat.* **2**, 081801 (2014).
40. Tikhomirov, G. et al. DNA-based programming of quantum dot valency, self-assembly and luminescence. *Nature Nanotechnol.* **6**, 485–490 (2011).
41. de Mello, J. C., Wittmann, H. F. & Friend, R. H. An improved experimental determination of external photoluminescence quantum efficiency. *Adv. Mater.* **9**, 230–232 (1997).
42. Manser, J. S. & Kamat, P. V. Band filling with free charge carriers in organometal halide perovskites. *Nature Photon.* **8**, 737–743 (2014).

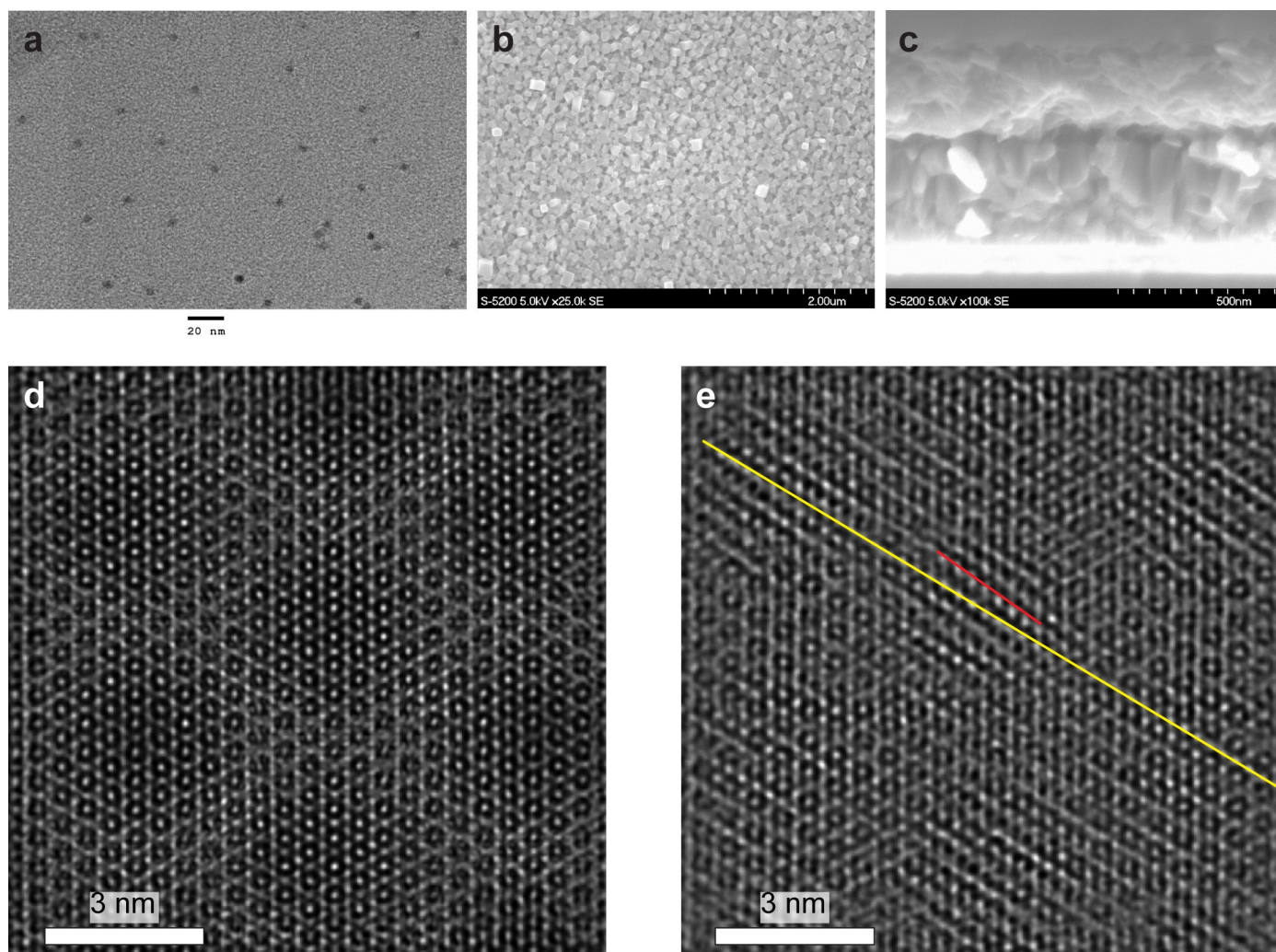


Extended Data Figure 1 | Planar averaged total charge difference with respect to the pristine slabs. The total charge difference ($\Delta\rho$) is measured along the Z axis of the material.



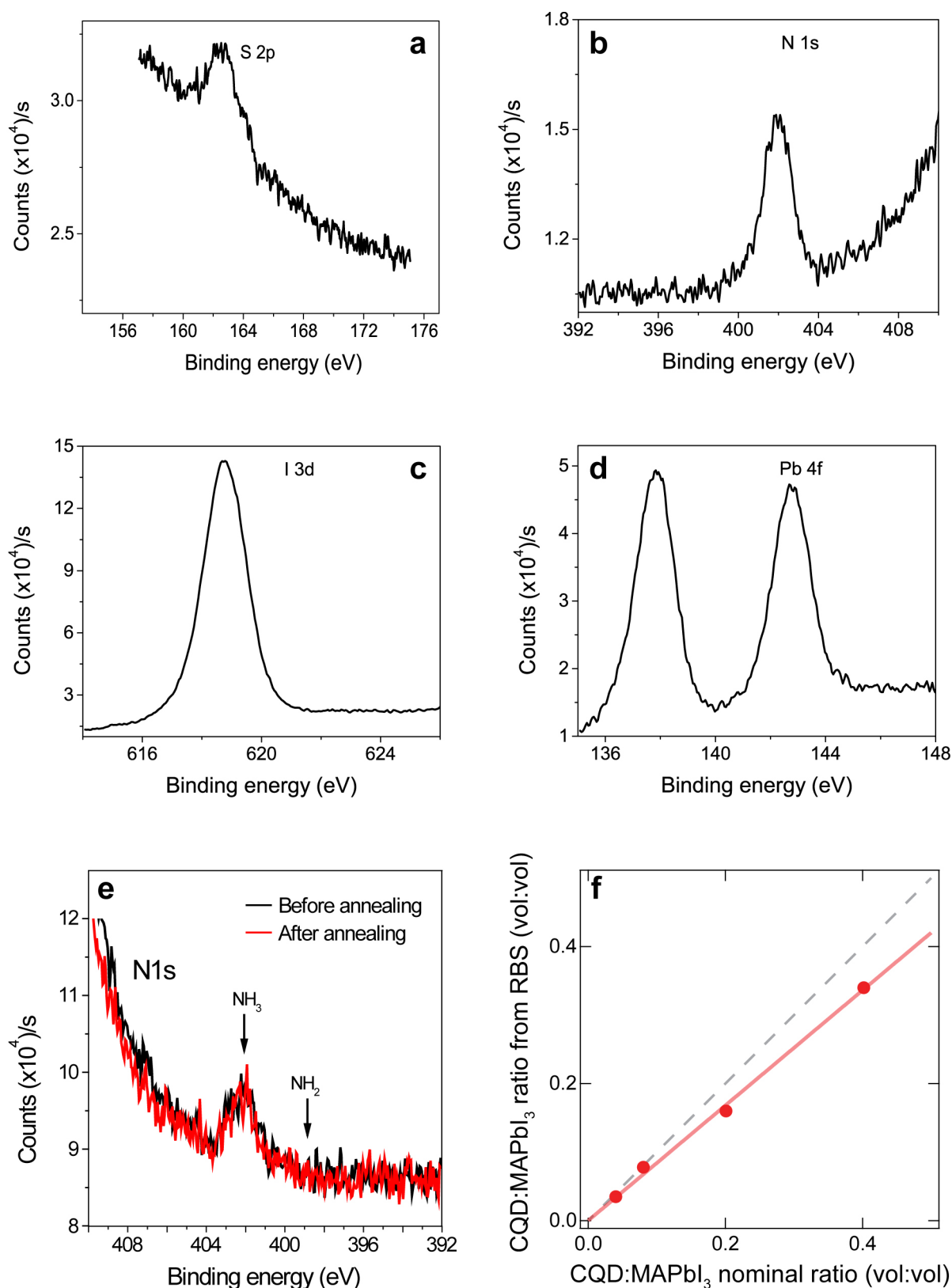
Extended Data Figure 2 | DFT simulation. **a, b**, Highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO), respectively, of a matched CQD. **c, d**, HOMO and LUMO, respectively, of mismatched dots. The states are mostly localized within the CQD in matched

dots, whereas they are localized on the interface between the CQDs and perovskite in mismatched dots, indicating the formation of defects in the latter. **e, f**, The projected density of states (PDOS) as a function of energy (E) of matched and mismatched (rotated) CQDs, respectively.



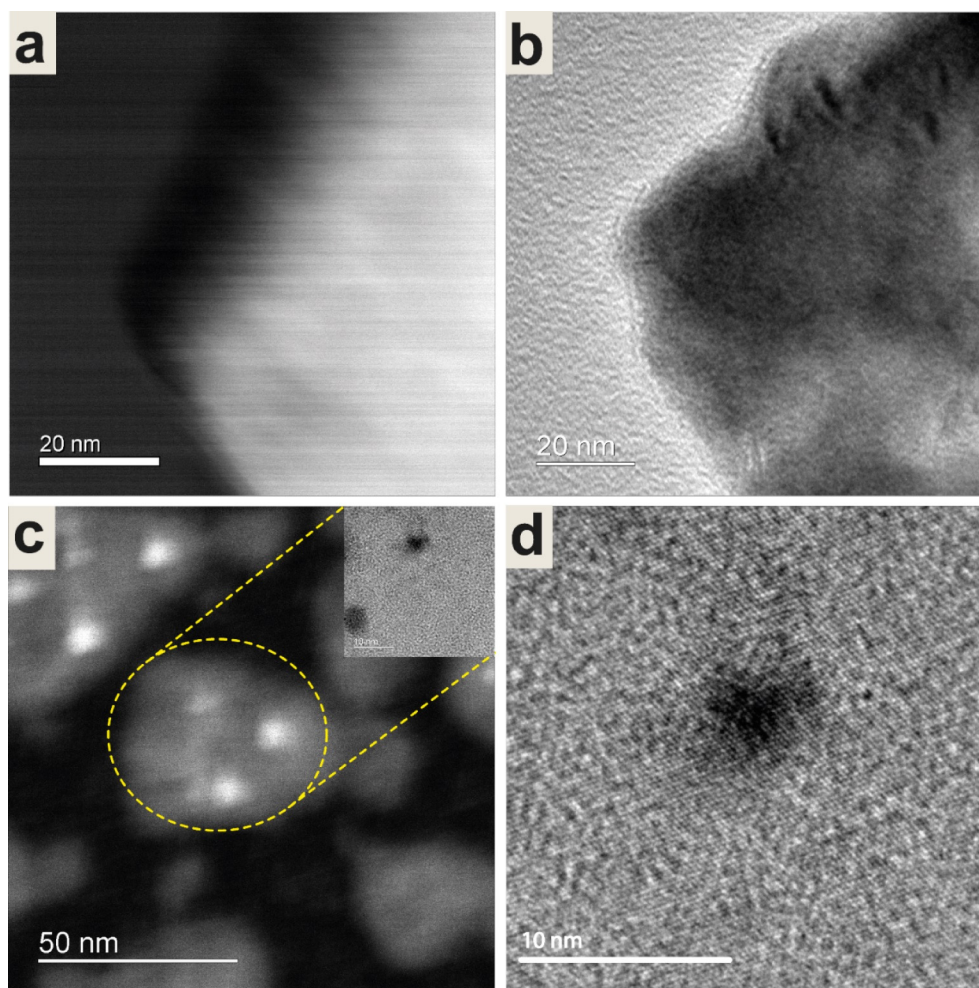
Extended Data Figure 3 | Microscopic images. **a**, TEM image of CQDs after solution-phase iodide–ligands exchange. The average CQD size is approximately 4 nm. **b**, Scanning electron microscopy (SEM) image of the film surface with CQDs embedded in perovskite matrix (CQD volume percentage of 3.8%). The average perovskite grain size is about 60 nm. **c**, SEM image of

the cross-section of the film with CQDs embedded in perovskite matrix (CQD volume percentage of 3.8%). The film thickness is about 240 nm (top layer). **d**, **e**, Simulated TEM images of matched (**d**) and mismatched (**e**) CQDs and perovskite. Yellow line, perovskite lattice plane; red line, CQD lattice plane.



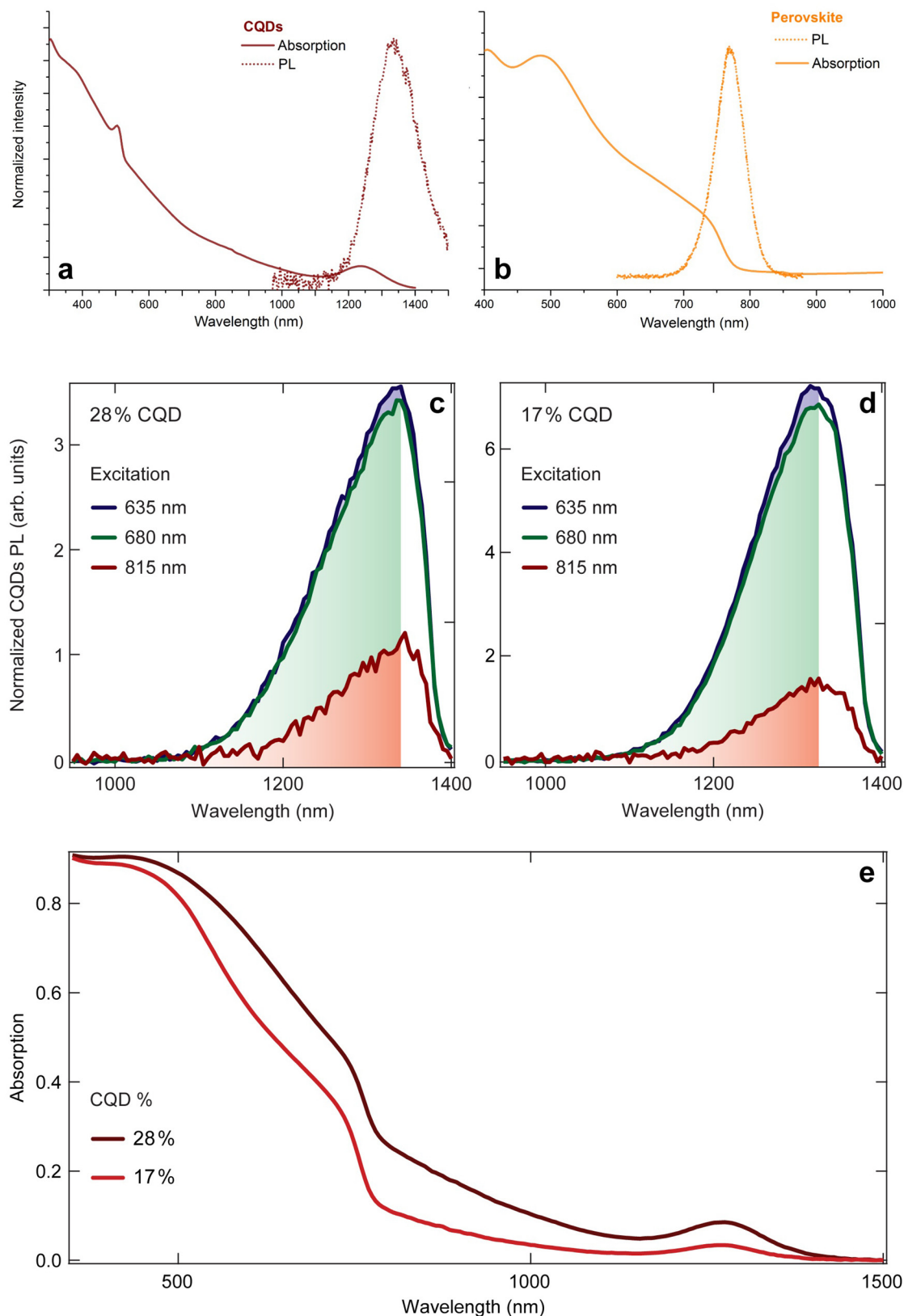
Extended Data Figure 4 | Compositional analysis of PHC. **a–d**, XPS analysis of the CQDs embedded in PHC film (Pb 4f, S 2p, N 1s, I 3d). All the elements (as labelled) are observed in the film, indicating the existence of both PbS and MAPbI₃. **e**, XPS spectra of the N 1s orbital for high-loading PbS films, which follows the ligand exchange to perovskite ligands in butylamine, before and after annealing at 70 °C for 10 minutes, before the methylammonium post-treatment step. We detect only the ammonium signal at 402 eV and no amine signal at 398–399 eV, which suggests that the methylammonium iodide ligand

is more stable on the surface and displaces weakly bound butylamine, even before film annealing. **f**, Plot of the nominal volume ratio of CQDs to perovskite versus the ratio measured using RBS (both with the perovskite value scaled to one), with a linear fit (red line) superimposed. The ideal case, where the measured ratio and the nominal ratio are equal, is indicated by the dashed grey line, showing the agreement between the nominal value and the one evaluated using RBS.



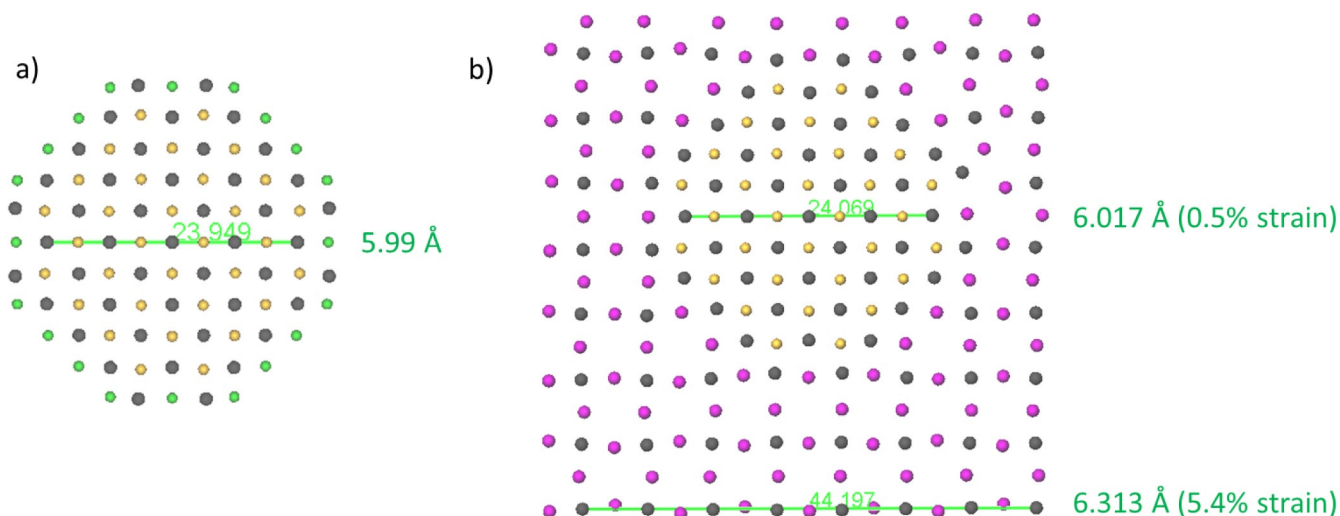
Extended Data Figure 5 | TEM images of PHC. **a, b**, HAADF-STEM and TEM images of pure perovskite, respectively. **c**, HAADF-STEM image of dots in perovskite. The region indicated by the dashed yellow circle contains three

PbS nanocrystals (white spots) located inside perovskite crystal (grey area). Inset, TEM image of part of this region. **d**, A close up of part of the inset in **c**.



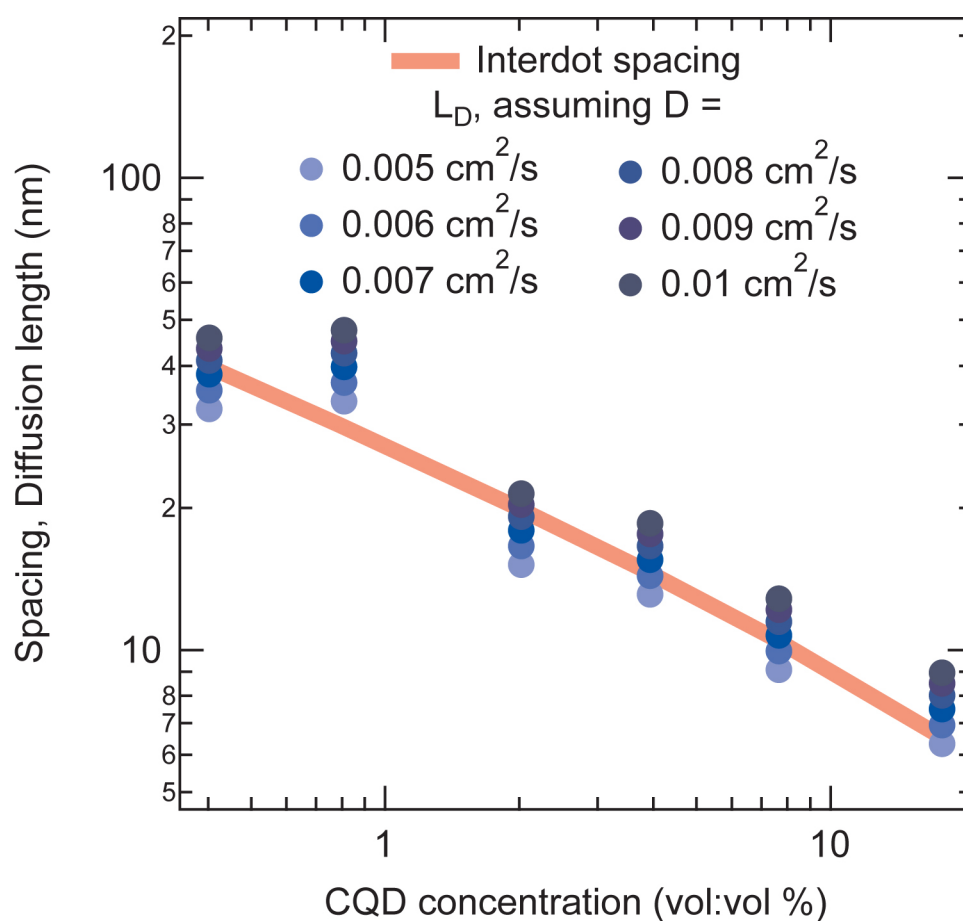
Extended Data Figure 6 | Absorption and photoluminescence properties of PHC. **a, b,** Absorption and photoluminescence (PL) spectra for a pure CQD film (**a**) and a pure perovskite film (**b**). **c, d,** CQD photoluminescence signal from PHC films with 28% (**c**) and 17% (**d**) CQD concentration, acquired at $\lambda = 635$ nm (blue), 680 nm (green), and 815 nm (red). The photoluminescence signal has been normalized by the excitation intensity at the different

wavelengths used. The shaded area corresponds the region of the spectrum used to calculate the photoluminescence integrated area in equation (3). **e,** Corresponding optical absorption spectra for a CQD concentration of 28% (dark red) and 17% (red) that were used to evaluate the absorption parameters in, for example, equation (3).



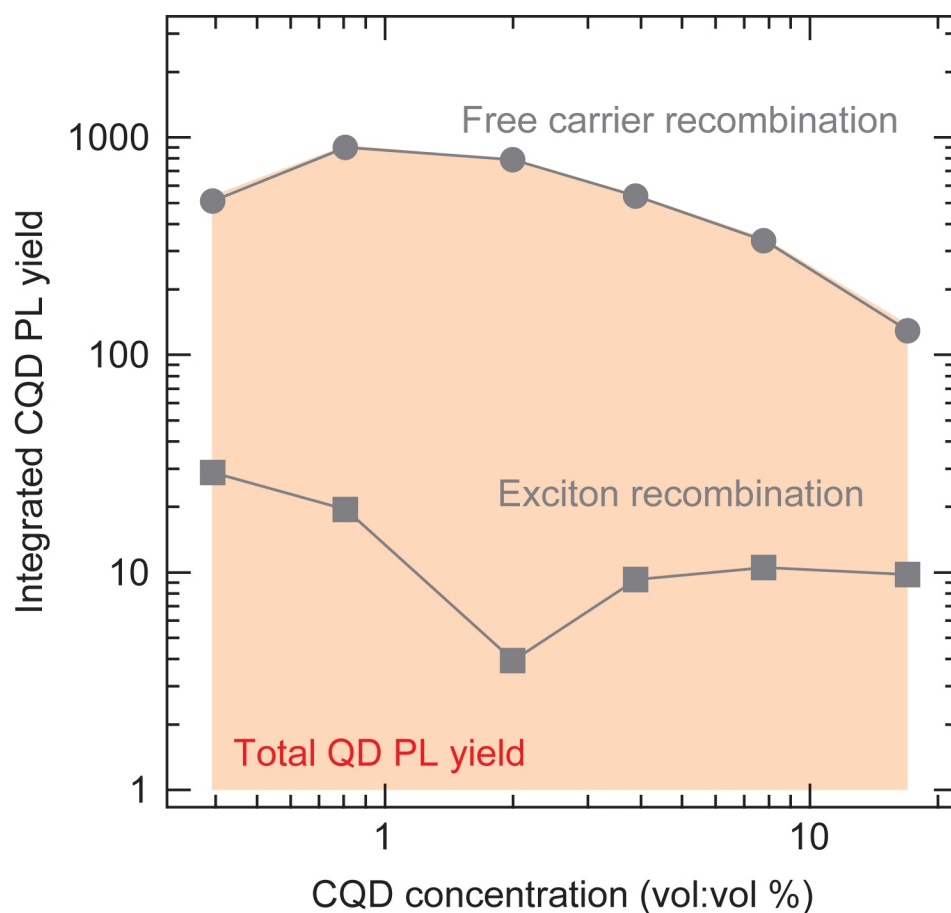
Extended Data Figure 7 | DFT simulation of strain distribution in PHC composite material. **a**, Cross-section of the relaxed 3-nm PbS CQD in vacuum. **b**, DFT-optimized geometry of the same 3-nm CQD epitaxially matched with perovskite in a unit cell corresponding to unstrained perovskite. Colours represent the following: grey, lead; yellow and green, sulphur;

pink, iodine. The green lines and corresponding numbers are the long-range Pb-to-Pb distance (23.949 Å, 24.069 Å and 44.197 Å); the large green numbers are the average lattice constant (average distance between lead atoms), along those lines.



Extended Data Figure 9 | CQD-concentration-dependent diffusion length. Plot of the diffusion length L_D of carriers in the perovskite matrix (round markers) for various CQD concentrations, calculated using equation (11) and

different possible values for the diffusion coefficient D , as labelled. The red line shows the calculated average inter-dot spacing as a function of CQD:MAPbI₃ volume ratio.



Extended Data Figure 10 | Integrated photoluminescence contributions of the CQDs. A stacked plot showing the total photoluminescence yield of the CQDs (shaded area), which is made up of contributions from free carrier recombination and exciton recombination: filled circles represent the integrated (with respect to time) photoluminescence (PL) contribution as a result of free carriers injected from the perovskite ('free carrier recombination'); filled squares

represent the integrated (with respect to time) photoluminescence as a result of the dissociation of excitons generated directly in the CQDs ('excitation recombination'); see also equation (10). Consistent with the photoluminescence excitation spectra, the largest contribution to the CQDs photoluminescence is ascribed to radiative recombination of free carriers transferred from the perovskite.

Extended Data Table 1 | Composition and photophysical parameters of quantum-dot-in-perovskite solid.

a

PbS	PbI ₂ : Mass:Mass ratio	MAPbI ₃ : PbS Vol:Vol ratio	PbS : MAPbI ₃ Vol:Vol ratio	PbS Volume percentage (%)	PbS : MAPbI ₃ Vol:Volume ratio from fit to RBS
	200	497.676	0.00201	0.20053	0.00169
	100	248.838	0.00402	0.40026	0.00338
	50	124.419	0.00804	0.79733	0.00676
	20	49.7676	0.02009	1.96976	0.01689
	10	24.8838	0.04019	3.86342	0.03378
	5	12.4419	0.08037	7.43942	0.06755
	4	9.95352	0.10047	9.12949	0.08444
	3	7.46514	0.13396	11.81315	0.11259
	2	4.97676	0.20093	16.73147	0.16888
	1	2.48838	0.40187	28.6666	0.33777

b

CQD Concentration	PL_QDs, short	PL_QDs long	I _{ex} , short	I _{ex} , long	A _{QDs} , long	A _{QDs} short	A _P , short	Transfer Efficiency
(vol:vol)	Counts	Counts	Counts	Counts				%
0.004002594	10553400	476450	2378155940	872728455	0.008	0.0220	0.351	12.55
0.007973274	6152760	448548	3203109586	1321620214	0.005	0.0103	0.396	4.97
0.019697602	11752660	838076	2378155940	872728455	0.011	0.0254	0.347	9.33
0.038634204	4893860	475011	3203109586	1321620214	0.015	0.0323	0.385	8.21
0.074394245	11586780	1175544	3203109586	1321620214	0.023	0.0492	0.368	12.03
0.091294853	13070740	1359321	3203109586	1321620214	0.024	0.0508	0.391	11.11
0.118131537	4870700	586011	3203109586	1321620214	0.080	0.1327	0.342	40.95
0.167314732	3940820	307583	3203109586	1321620214	0.067	0.1126	0.355	68.42
0.286666017	2239240	256204	3203109586	1321620214	0.164	0.2652	0.407	80.37

a, List of the values for the nominal MAPbI₃ to PbS mass ratio (first column, PbS mass scaled to one) and volume ratio (second column, PbS volume scaled to one), PbS to MAPbI₃ volume ratio (third column, reciprocal of second column), and PbS volume as a percentage of the total volume (fourth column). The experimental values for the PbS volume as a percentage of the total volume, as extrapolated from a fit to the RBS data (see also Extended Data Fig. 4) are shown in the fifth column. **b**, Numerical values of the photophysical parameters in equations (1)–(3) that lead to the evaluation of the total (carrier) transfer efficiency η_{tot} .

Nanotubes mediate niche–stem–cell signalling in the *Drosophila* testis

Mayu Inaba^{1,2,3}, Michael Buszczak³ & Yukiko M. Yamashita^{1,2}

Stem cell niches provide resident stem cells with signals that specify their identity. Niche signals act over a short range such that only stem cells but not their differentiating progeny receive the self-renewing signals¹. However, the cellular mechanisms that limit niche signalling to stem cells remain poorly understood. Here we show that the *Drosophila* male germline stem cells form previously unrecognized structures, microtubule-based nanotubes, which extend into the hub, a major niche component. Microtubule-based nanotubes are observed specifically within germline stem cell populations, and require intraflagellar transport proteins for their formation. The bone morphogenetic protein (BMP) receptor Tkv localizes to microtubule-based nanotubes. Perturbation of microtubule-based nanotubes compromises activation of Dpp signalling within germline stem cells, leading to germline stem cell loss. Moreover, Dpp ligand and Tkv receptor interaction is necessary and sufficient for microtubule-based nanotube formation. We propose that microtubule-based nanotubes provide a novel mechanism for selective receptor–ligand interaction, contributing to the short-range nature of niche–stem–cell signalling.

The *Drosophila* testis represents an excellent model system to study niche–stem–cell interactions because of its well-defined anatomy: eight to ten germline stem cells (GSCs) are attached to a cluster of somatic hub cells, which serve as a major component of the stem cell niche

(Fig. 1a). The hub secretes at least two ligands: the cytokine-like ligand Unpaired (Upd), and a BMP ligand Decapentaplegic (Dpp), both of which regulate GSC maintenance^{2–5}. GSCs typically divide asymmetrically, so that one daughter of the stem cell division remains attached to the hub and retains stem cell identity, while the other daughter, called a gonialblast, is displaced away from the hub and initiates differentiation⁶. Given the close proximity of GSCs and gonialblasts, the ligands (Upd and Dpp) must act over a short range so that signalling is only active in stem cells, but not in differentiating germ cells. The basis for this sharp boundary of pathway activation remains poorly understood.

Using green fluorescent protein (GFP)- α 1-tubulin^{84B} expressed in germ cells (*nos-gal4*>*UAS-GFP- α tub*), we found that GSCs form protrusions, referred to as microtubule-based (MT)-nanotubes hereafter, that extend into the hub (Fig. 1b). MT-nanotubes are sensitive to fixation similar to other thin protrusions reported so far, such as tunnelling nanotubes⁷ and cytonemes⁸, explaining why they have escaped detection in previous studies. MT-nanotubes appear to be specific to GSCs: we observed 6.67 MT-nanotubes per testis in the GSC population (or 0.82 per cell, $n = 73$ testes). The average thickness and length of MT-nanotubes are $0.43 \pm 0.29 \mu\text{m}$ (at the base of MT-nanotube, $n = 51$ nanotubes) and $3.32 \pm 1.6 \mu\text{m}$ ($n = 82$ nanotubes), respectively. These GSC MT-nanotubes are uniformly oriented towards the hub area (Fig. 1c). By contrast, differentiating germ cells

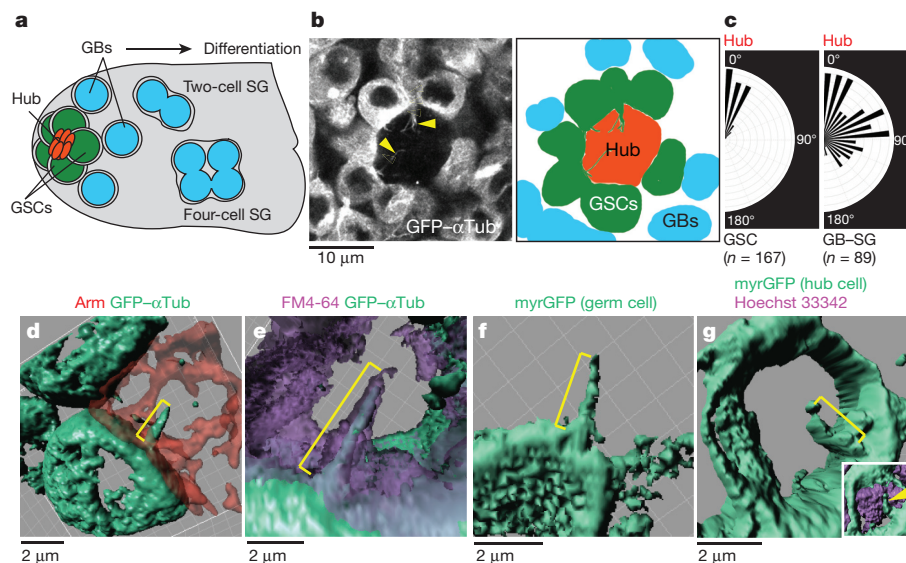


Figure 1 | Characterization of MT-nanotubes in *Drosophila* male GSC niche. **a**, Schematic of the *Drosophila* male GSC niche. GSCs are attached to the hub cells. The immediate daughters of GSCs, the gonialblasts (GBs) are displaced away from the hub, and become spermatogonia (SG). **b**, An apical tip of the testis expressing GFP- α Tub in germ cells (*nos-gal4*>*GFP- α tub*). MT-nanotubes are indicated by arrowheads. A graphic interpretation is shown in the right-hand panel. **c**, Orientation of nanotubes towards the hub

in GSCs versus gonialblasts/spermatogonia. The size of each vector represents the frequency of MT-nanotubes oriented towards each direction. Indicated numbers of nanotubes ($n > 30$ testes) were scored from three independent experiments. **d–g**, Three-dimensional rendering images of MT-nanotubes (brackets) in fixed (**d**) or live tissue (**e–g**), with indicated cell membrane markers; *nos-gal4*>*GFP- α tub* was used for **d** and **e**.

¹Life Sciences Institute, Department of Cell and Developmental Biology Medical School, University of Michigan, Ann Arbor, Michigan 48109, USA. ²Howard Hughes Medical Institute, University of Michigan Ann Arbor, Michigan 48109, USA. ³Department of Molecular Biology, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75390, USA.

showed only 0.44 MT-nanotubes per testis (or <0.002 per cell, $n = 75$ testes), without any particular orientation when present (Fig. 1c). MT-nanotubes were sensitive to colcemid, the microtubule-depolymerizing drug, but not to the actin polymerization inhibitor cytochalasin B, suggesting that MT-nanotubes are microtubule-based structures (Extended Data Fig. 1a–d, f). MT-nanotubes were not observed in mitotic GSCs (Extended Data Fig. 1e, g), and GSCs form new MT-nanotubes as they exit from mitosis (Extended Data Fig. 1h and Supplementary Video 1). By contrast, MT-nanotubes in interphase GSCs were stably maintained for up to 1 h of time-lapse live imaging (Supplementary Video 2). Although cell-cycle-dependent formation of MT-nanotube resembles that of primary cilia^{9,10}, MT-nanotubes are distinct structures, in that they lack acetylated microtubules and are sensitive to fixation. Furthermore, a considerable fraction of GSCs form multiple MT-nanotubes per cell (54% of GSCs with MT-nanotubes, $n = 251$ GSCs), and MT-nanotubes are not always associated with the centrosome/basal body, as is the case for the primary cilia (Extended Data Fig. 1i).

To examine the geometric relationship between MT-nanotubes and hub cells further, we imaged MT-nanotubes in combination with various cell membrane markers, followed by three-dimensional rendering. Although the MT-nanotubes are best visualized in unfixed testes that express GFP- α Tub in germ cells, adding a low concentration (1 μ M) of taxol to the fixative preserves MT-nanotubes, allowing immunofluorescence staining. First, Armadillo (Arm, β -catenin) staining, which marks adherens junctions formed at hub cell/hub cell as well as hub cell/GSC boundaries, revealed that adherens junctions do not form on the surface of MT-nanotubes (Fig. 1d and Supplementary Video 3). Using FM4-64 styryl dye, we found that the MT-nanotubes are ensheathed by membrane lipids (Fig. 1e and Supplementary Videos 4 and 5). Furthermore, myristoylation/palmitoylation site GFP (myrGFP), a membrane marker, expressed in either the germline (Fig. 1f) or hub cells (Fig. 1g) illuminated MT-nanotubes, suggesting that the surface membrane of a MT-nanotube is juxtaposed to hub-cell plasma membrane.

We examined genes that regulate primary cilia and cytonemes for their possible involvement in MT-nanotube formation (Fig. 2a). RNA interference (RNAi)-mediated knockdown of *oseg2* (IFT172), *osm6* (IFT52) and *che-13* (IFT57), components of the intraflagellar transport (IFT)-B complex that are required for primary cilium anterograde transport and assembly¹¹, significantly reduced the length and the frequency of MT-nanotubes (Fig. 2a, Extended Data Fig. 2b and Extended Data Table 1). Knockdown of *Dlic*, a dynein intermediate chain required for retrograde transport in primary cilia¹², also reduced the MT-nanotube length and frequency (Fig. 2a and Extended Data Table 1). Knockdown of *kpl10A*, a *Drosophila* homologue of mammalian *kif24* (a MT-depolymerizing kinesin of the kinesin-13 family, which suppresses precocious cilia formation¹³), resulted in abnormally thick/bulged MT-nanotubes (Fig. 2a, Extended Data Fig. 2c and Extended Data Table 1). We did not observe significant changes in MT-nanotube morphology upon knockdown of IFT-A retrograde transport genes, such as *oseg1* and *oseg3* (Fig. 2a and Extended Data Table 1).

Endogenous Klp10A localized to MT-nanotubes both in wild-type testes and in GFP- α Tub-expressing testes (Fig. 2b and Extended Data Fig. 2d, e). GFP-*Oseg2* (IFT-B), GFP-*Oseg1*, GFP-*Oseg3* (IFT-A) and *Dlic* also localized to the MT-nanotubes when expressed in germ cells (Fig. 2c and Extended Data Fig. 2f–i). The localization of IFT-A components to MT-nanotubes, without detectable morphological abnormality upon mutation/knockdown, is reminiscent of the observation that most of the genes for IFT-A are not required for primary cilia assembly^{14–17}. Expression of a dominant negative form of *Dia* (*Dia*^{DN}) or a temperature-sensitive form of *Shi* (*Shi*^{ts}) in germ cells (*nos-gal4*>*UAS-dia*^{DN} or *UAS-shi*^{ts}), which perturb cytoneme formation¹⁸, did not influence the morphology or frequency of MT-nanotubes in GSCs (Fig. 2a and Extended Data Table 1).

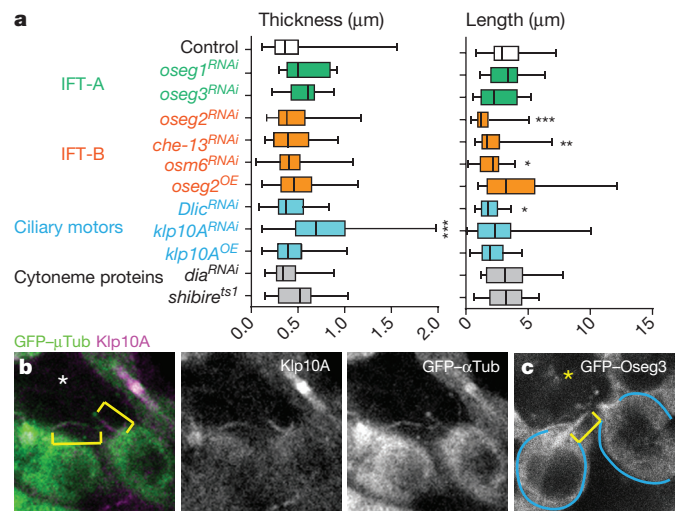


Figure 2 | IFT genes are required for MT-nanotube formation. **a**, Effect of RNAi-mediated knockdown or overexpression (OE) of indicated genes on MT-nanotube morphology. Box plot shows 25–75% (box), median (band inside) and minima to maxima (whiskers). Indicated numbers of MT-nanotubes (Extended Data Table 1) from at least two independent crosses were scored for each data point. *P* values from *t*-tests are provided as **P* ≤ 0.05, ***P* ≤ 0.01 and ****P* ≤ 0.001. **b**, Examples of MT-nanotubes stained by anti-Klp10A antibody in GFP- α Tub-expressing testes. **c**, Apical testis tip expressing GFP-*Oseg3* in germ cells. MT-nanotube is indicated by brackets. GSCs are indicated by blue lines. Asterisk indicates hub. Scale bar, 10 μ m.

Taken together, these results show that primary cilia proteins localize to MT-nanotubes and regulate their formation.

In search of the possible involvement of MT-nanotubes in hub–GSC signalling, we found that the Dpp receptor, Thickveins (Tkv), expressed in germ cells (*nos-gal4*>*tkv-GFP*) was observed within the hub region (Extended Data Fig. 3a), in contrast to GFP alone, which remained within the germ cells (Extended Data Fig. 3b). A GFP protein trap of Tkv (in which GFP tags Tkv at the endogenous locus) also showed the same localization pattern as Tkv-GFP expressed by *nos-gal4* (Extended Data Fig. 3c). By inducing GSC clones that co-express Tkv-mCherry and GFP- α Tub, we found that Tkv-mCherry localizes along the MT-nanotubes as puncta (Fig. 3a). Furthermore, using live observation, Tkv-mCherry puncta were observed to move along the MT-nanotubes marked with GFP- α Tub (Extended Data Fig. 3d), suggesting that Tkv is transported towards the hub along the MT-nanotubes. It should be noted that, in the course of our study, we noticed that mCherry itself localized to the hub when expressed in germ cells, similar to Tkv-GFP and Tkv-mCherry (see Extended Data Fig. 3e, f and Supplementary Note 1). Importantly, the receptor for Upd, Domeless (Dome), predominantly stayed in the cell body of GSCs (Extended Data Fig. 3g), demonstrating the specificity/selectivity of MT-nanotubes in trafficking specific components of the niche signalling pathways. A reporter of ligand-bound Tkv, TIPF¹⁹, localized to the hub region together with Tkv-mCherry (Fig. 3b), in addition to its reported localization at the hub–GSC interface¹⁹. Furthermore, Dpp-GFP expressed by hub cells co-localized with Tkv-mCherry expressed in germline (Fig. 3c, *dpp-lexA*^{ts}>*dpp-GFP*, *nos-gal4*^{ts}>*tkv-mCherry*). These results suggest that ligand (Dpp)-receptor (Tkv) engagement and activation occurs at the interface of the MT-nanotube surface and the hub cell plasma membrane. Knockdown of IFT-B components (*oseg2*^{RNAi}, *che-13*^{RNAi} or *osm6*^{RNAi}), which reduces MT-nanotube formation, resulted in reduction of the number of Tkv-GFP puncta in the hub area, concomitant with increased membrane localization of Tkv-GFP (Fig. 3d, f, g). A similar trend was observed upon treatment of the testes with colcemid (Extended Data Fig. 3h, i), suggesting that MT-nanotubes are required for trafficking of Tkv into the hub area. By contrast, knockdown of Klp10A, which causes

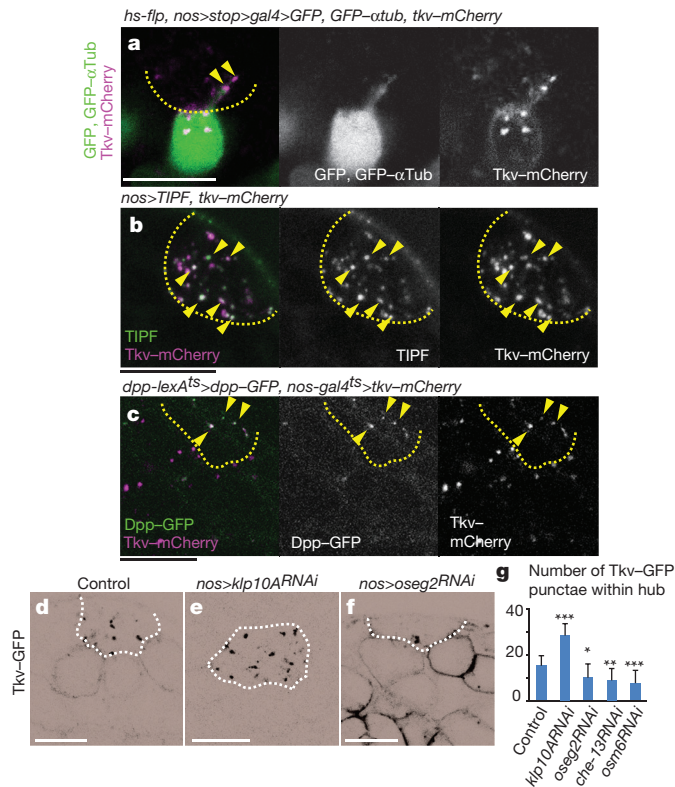


Figure 3 | Dpp signalling components localize to the MT-nanotubes. **a**, A GSC clone expressing TkV-mCherry, GFP- α Tub and GFP (*hs-flp, nos>stop>gal4>GFP, GFP- α tub, tkv-mCherry*). **b**, An apical tip of the testis expressing TIPF and TkV-mCherry in germ cells. Arrowheads point to a few of co-localizing puncta. **c**, An apical tip of the testis expressing Dpp in the hub and TkV in germ cells (*dpp-lexA^{ts}>dpp-GFP, nos-gal4^{ts}>tkv-mCherry*). **d–f**, TkV-GFP expressed in control (**d**), *kfp10A^{RNAi}* (**e**) and *oseg2^{RNAi}* (**f**) germ cells (*nos-gal4^{ts}>UAS-tkv-GFP, UAS-RNAi*). Black and white of micrograph were inverted for better visibility of TkV localization to the hub and plasma membrane. **g**, Average number and standard deviations of TkV-GFP puncta within hub area per testis for indicated genotypes. $n = 15$ testes from at least two independent crosses were scored. P values from t -tests are provided as $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$. Scale bar, 10 μ m.

thickening of MT-nanotubes, led to an increase in the number of TkV-GFP puncta in the hub area (Fig. 3d, e, g). Taken together, these data suggest that TkV is trafficked into the hub via MT-nanotubes, where it interacts with Dpp secreted from the hub.

Knockdown of *kfp10A* (*kfp10A^{RNAi}*) led to elevated phosphorylated Mad (pMad) levels, a readout of Dpp pathway activation, in GSCs (Fig. 4a, b, d and Supplementary Note 2). By contrast, RNAi-mediated knockdown of *oseg2*, *osm6* and *che-13* (IFT-B components), which causes shortening of MT-nanotubes, reduced the levels of pMad in GSCs (Fig. 4c, d). Dad-LacZ, another readout of Dpp signalling activation, exhibited clear upregulation upon knockdown of *kfp10A* (Extended Data Fig. 4a, b). GSC clones of *che-13^{RNAi}*, *osm6^{RNAi}* or *oseg2⁴⁵²* were lost rapidly compared with control clones (Fig. 4e, f), consistent with the idea that MT-nanotubes help to promote Dpp signal transduction^{3,4}. Knockdown of *oseg2*, *che-13* and *osm6* did not visibly affect cytoplasmic microtubules (Extended Data Fig. 4d–g), suggesting that GSC maintenance defects upon knockdown of these genes are probably mediated by their role in MT-nanotube formation. Global RNAi knockdown of these genes in all GSCs using *nos-gal4* did not cause a significant decrease in GSC numbers (data not shown), indicating that compromised Dpp signalling due to MT-nanotube reduction leads to a competitive disadvantage in regards to GSC maintenance only when surrounded by wild-type GSCs.

When *kfp10A^{RNAi}* GSC clones were induced, pMad levels specifically increased in those GSC clones, indicating that Kfp10A acts

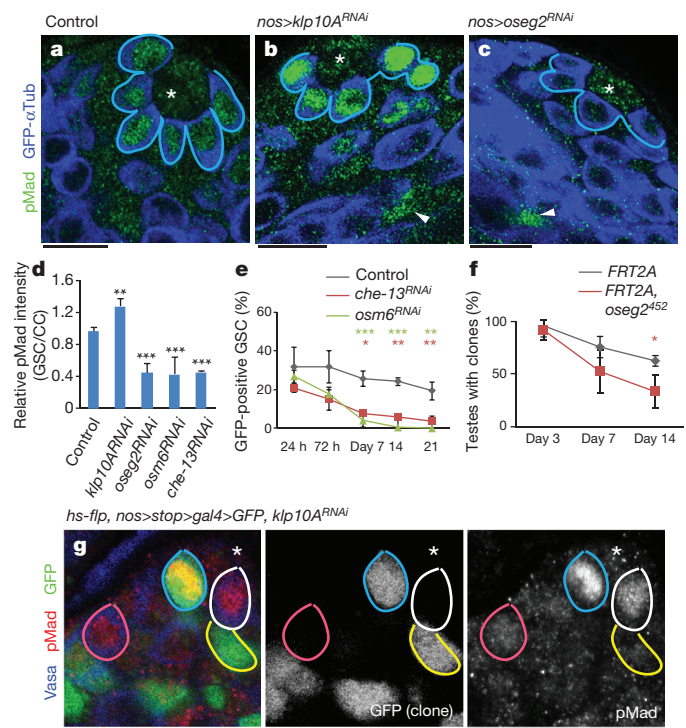


Figure 4 | MT-nanotubes are required for Dpp signalling activation and GSC maintenance. **a–c**, pMad staining in control (**a**), *kfp10A^{RNAi}* (**b**) and *oseg2^{RNAi}* (**c**) testes. pMad signal in somatic cyst cells (arrowheads), which remains unaffected by germ-cell specific modulation of MT-nanotube components, was used to normalize pMad levels in GSCs. **d**, Quantification of pMad intensity in the GSCs of indicated genotypes. Indicated numbers of GSCs (Extended Data Table 1) from at least two independent crosses were scored for each data point. **e**, **f**, Maintenance of *che-13^{RNAi}*, *osm6^{RNAi}* (**e**) and *oseg2⁴⁵²* (**f**) mutant GSC clones. Indicated numbers of GSCs (Supplementary Table 1) from at least two independent experiments were scored for each data point. **g**, A *kfp10A^{RNAi}* GSC clone (72 h after clone induction, blue circle) with a higher pMad level, compared with control GSCs (white circle), *kfp10A^{RNAi}* spermatogonia clone (yellow circle) and control spermatogonia clone (pink circle) have similar pMad levels. Asterisk indicates hub. Scale bar, 10 μ m. Average value and standard deviations are shown in each graph. P values from t -tests are provided as $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$.

cell-autonomously in GSCs to influence Dpp signal transduction (Fig. 4g). Importantly, *kfp10A^{RNAi}* spermatogonia (Fig. 4g, yellow line) did not show a significant elevation in pMad level compared with control spermatogonia (Fig. 4g, pink line), demonstrating that the role of Kfp10A in regulation of Dpp pathway is specific to GSCs. pMad levels did not change in spermatogonia upon manipulation of MT-nanotube formation (Extended Data Fig. 4c). GSC clones of *kfp10A^{RNAi}* or *kfp10A* null mutant (*kfp10A²⁴*) did not dominate in the niche, despite upregulation of pMad (Extended Data Fig. 5), possibly because of its known role in mitosis²⁰. Importantly, these conditions did not significantly change STAT92E levels, which reflect Upd-JAK-STAT signalling in GSCs^{2,21}, revealing the selective requirement of MT-nanotubes in Dpp signalling (Extended Data Fig. 6). Together, these results demonstrate that MT-nanotubes specifically promote Dpp signalling and their role in enhancing the Dpp pathway is GSC specific.

Since cytonemes are induced/stabilized by the signalling molecules themselves¹⁸, we explored the possible involvement of Dpp in MT-nanotube formation. First, we found that a temperature-sensitive *dpp* mutant (*dpp^{hr56}/dpp^{hr4}*) exhibited a dramatic decrease in the frequency of MT-nanotubes (0.067 MT-nanotubes per GSC, $n = 244$ GSCs) and the remaining MT-nanotubes were significantly thinner (Fig. 5a, b, Extended Data Fig. 7a, b and Extended Data Table 1). Knockdown of *tkv* (*tkv^{RNAi}*) in GSCs also resulted in reduced length and frequency of MT-nanotubes (Fig. 5a, b, Extended Data Fig. 7c and Extended Data

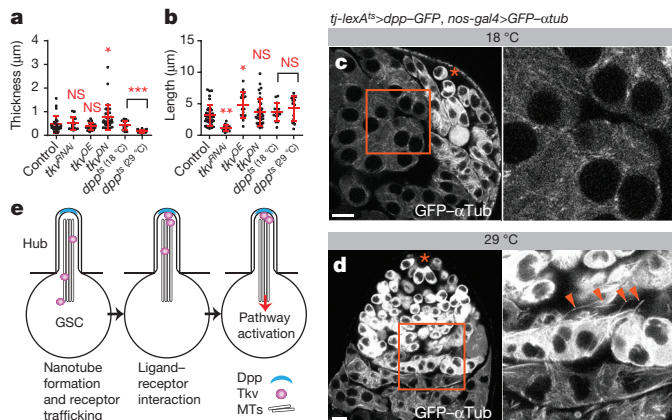


Figure 5 | Dpp signalling is necessary and sufficient for MT-nanotube formation. **a, b**, Quantification of MT-nanotube thickness and length in GSCs of indicated genotypes. Each scored value is plotted as a dot. Red line indicates average value and standard deviations. Indicated numbers of nanotubes (Extended Data Table 1) from at least two independent crosses were scored for each genotype. *P* values from *t*-test are provided as **P* ≤ 0.05, ***P* ≤ 0.01, ****P* ≤ 0.001; NS, non-significant (*P* > 0.05). **c, d**, MT-nanotube formation in absence (**c**) or presence (**d**) of Dpp expression in somatic cyst cells. Magnified images of squared regions are shown in right-hand panels. Arrowheads indicate ectopic MT-nanotubes. Asterisk indicates hub. Scale bar, 10 μm. **e**, Model. Dpp induces MT-nanotube formation, and receptor–ligand interaction occurs at the surface of MT-nanotubes, leading to signalling activation in GSCs.

Table 1). Conversely, overexpression of Tkv (*tkv^{OE}*)²² in germ cells led to significantly longer MT-nanotubes (Fig. 5a, b, Extended Data Fig. 7d and Extended Data Table 1). Interestingly, expression of a dominant negative Tkv (*tkv^{DN}*), which has intact ligand-binding domain but lacks its intracellular GS domain and kinase domain, resulted in thickening of MT-nanotubes, rather than reducing the thickness/length (Fig. 5a, b and Extended Data Table 1). This indicates that ligand–receptor interaction, but not downstream signalling events, is sufficient to induce MT-nanotube formation. Strikingly, upon ectopic expression of Dpp in somatic cyst cells (*tj-lexA>dpp*), spermatogonia/spermatocytes were observed to have numerous MT-nanotubes (Fig. 5c, d and Extended Data Fig. 7e), suggesting that Dpp is necessary and sufficient to induce or stabilize MT-nanotubes in the neighbouring germ cells. In turn, MT-nanotubes may promote selective ligand–receptor interaction between hub and GSCs, leading to spatially confined self-renewal (Fig. 5e).

Our study shows that previously unrecognized structures, MT-nanotubes, extend into the hub to mediate Dpp signalling. We propose that MT-nanotubes form a specialized cell surface area, where productive ligand–receptor interaction occurs. In this manner, only GSCs can access the source of highest ligand concentration in the niche via MT-nanotubes, whereas gonialblasts do not experience the threshold of signal transduction necessary for self-renewal, contributing to the short-range nature of niche signalling. In summary, the results reported here illuminate a novel mechanism by which the niche specifies stem cell identity in a highly selective manner.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 October 2014; accepted 1 June 2015.

Published online 1 July 2015.

1. Morrison, S. J. & Spradling, A. C. Stem cells and niches: mechanisms that promote stem cell maintenance throughout life. *Cell* **132**, 598–611 (2008).

2. Tulina, N. & Matunis, E. Control of stem cell self-renewal in *Drosophila* spermatogenesis by JAK-STAT signaling. *Science* **294**, 2546–2549 (2001).
3. Kiger, A. A., Jones, D. L., Schulz, C., Rogers, M. B. & Fuller, M. T. Stem cell self-renewal specified by JAK-STAT activation in response to a support cell cue. *Science* **294**, 2542–2545 (2001).
4. Shivdasani, A. A. & Ingham, P. W. Regulation of stem cell maintenance and transit amplifying cell proliferation by TGF-β signaling in *Drosophila* spermatogenesis. *Curr. Biol.* **13**, 2065–2072 (2003).
5. Kawase, E., Wong, M. D., Ding, B. C. & Xie, T. Gbb/Bmp signaling is essential for maintaining germline stem cells and for repressing bam transcription in the *Drosophila* testis. *Development* **131**, 1365–1375 (2004).
6. Yamashita, Y. M., Jones, D. L. & Fuller, M. T. Orientation of asymmetric stem cell division by the APC tumor suppressor and centrosome. *Science* **301**, 1547–1550 (2003).
7. Davis, D. M. & Sowinski, S. Membrane nanotubes: dynamic long-distance connections between animal cells. *Nature Rev. Mol. Cell Biol.* **9**, 431–436 (2008).
8. Ramírez-Weber, F. A. & Kornberg, T. B. Cytonemes: cellular processes that project to the principal signaling center in *Drosophila* imaginal discs. *Cell* **97**, 599–607 (1999).
9. Avasthi, P. & Marshall, W. F. Stages of ciliogenesis and regulation of ciliary length. *Differentiation* **83**, S30–S42 (2012).
10. Pedersen, L. B. & Rosenbaum, J. L. Intraflagellar transport (IFT) role in ciliary assembly, resorption and signalling. *Curr. Top. Dev. Biol.* **85**, 23–61 (2008).
11. Goetz, S. C. & Anderson, K. V. The primary cilium: a signalling centre during vertebrate development. *Nature Rev. Genet.* **11**, 331–344 (2010).
12. Perrone, C. A. et al. A novel dynein light intermediate chain colocalizes with the retrograde motor for intraflagellar transport at sites of axoneme assembly in *Chlamydomonas* and mammalian cells. *Mol. Biol. Cell* **14**, 2041–2056 (2003).
13. Kobayashi, T., Tsang, W. Y., Li, J., Lane, W. & Dynlacht, B. D. Centriolar kinesin Kif24 interacts with CP110 to remodel microtubules and regulate ciliogenesis. *Cell* **145**, 914–925 (2011).
14. Iomini, C., Li, L., Esparza, J. M. & Dutcher, S. K. Retrograde intraflagellar transport mutants identify complex A proteins with multiple genetic interactions in *Chlamydomonas reinhardtii*. *Genetics* **183**, 885–896 (2009).
15. Tran, P. V. et al. THM1 negatively modulates mouse sonic hedgehog signal transduction and affects retrograde intraflagellar transport in cilia. *Nature Genet.* **40**, 403–410 (2008).
16. Qin, J., Lin, Y., Norman, R. X., Ko, H. W. & Eggenschwiler, J. T. Intraflagellar transport protein 122 antagonizes Sonic Hedgehog signaling and controls ciliary localization of pathway components. *Proc. Natl Acad. Sci. USA* **108**, 1456–1461 (2011).
17. Cortellino, S. et al. Defective ciliogenesis, embryonic lethality and severe impairment of the Sonic Hedgehog pathway caused by inactivation of the mouse complex A intraflagellar transport gene *Ift122/Wdr10*, partially overlapping with the DNA repair gene *Med1/Mbd4*. *Dev. Biol.* **325**, 225–237 (2009).
18. Roy, S., Huang, H., Liu, S. & Kornberg, T. B. Cytoneme-mediated contact-dependent transport of the *Drosophila* decapentaplegic signaling protein. *Science* **343**, 1244624 (2014).
19. Michel, M., Raabe, I., Kupinski, A. P., Perez-Palencia, R. & Bokel, C. Local BMP receptor activation at adherens junctions in the *Drosophila* germline stem cell niche. *Nature Commun.* **2**, 415 (2011).
20. Goshima, G. & Vale, R. D. Cell cycle-dependent dynamics and regulation of mitotic kinesins in *Drosophila* S2 cells. *Mol. Biol. Cell* **16**, 3896–3907 (2005).
21. Schulz, C. et al. A misexpression screen reveals effects of bag-of-marbles and TGFβ class signaling on the *Drosophila* male germ-line stem cell lineage. *Genetics* **167**, 707–723 (2004).
22. Dudu, V. et al. Postsynaptic Mad signaling at the *Drosophila* neuromuscular junction. *Curr. Biol.* **16**, 625–635 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Kornberg, S. Roy, T. Avidor-Reiss, E. Laufer, A. Rodal, D. Sharp, S. Noselli, A. C. Spradling, T. E. Haerry, E. R. Gavis, C.-Y. Lee, T. Xie, B. McCabe, K. S. McKim, Bloomington *Drosophila* Stock Center, Vienna *Drosophila* Resource Center and the Developmental Studies Hybridoma Bank for reagents; S. Roy, T. Kornberg, G. Boekhoff-Falk and D. King for comments and advice; K. Luby-Phelps, A. Bugde and M. Acar for advice for imaging/image data processing; and the Yamashita and Buszczak laboratory members for discussion. The research in the Yamashita laboratory is supported by the Howard Hughes Medical Institute. Y.M.Y. is supported by the MacArthur Foundation.

Author Contributions M.I. conceived the project, and executed experiments. All authors designed experiments, analysed the data, and wrote and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B. (michael.buszczak@utsouthwestern.edu) or Y.M.Y. (yukikomy@umich.edu).

METHODS

Fly husbandry and strains. All fly stocks were raised on standard Bloomington medium. The following fly stocks were used: *10XUAS-IVS-myr::GFP* (BDSC32197), *upd-gal4* (BDSC26796), *UAS-GFP-xtub* (BDSC7253 or BDSC7373), *UAS-dpp-GFP* (BDSC53716), *UAS-mCherry* (BDSC35787) and *hs-bam* (BDSC24636) were obtained from Bloomington Stock Center. *UAS-TIPF¹⁹*, *UAS-*tkv*-GFP*, *UAS-*tkv*-mCherry¹⁸*, *dpp-lexA* (LHG)²³, *lexAop-dpp²³*, *UAS-shi^{ts}* (ref. 18), *UAS-did^{RNAi}* (ref. 18) and *UAS-*tkv*^{DN}2dGSK-3D2* (ref. 18) were gifts from T. Kornberg and S. Roy; *dad-lacZ* (FBti0009617) was a gift from T. Xie; *UAS-*tkv*-GFP* transgene²² was a gift from A. Rodal; *tkv-GFP* protein trap line (CPTI-002487) was a gift from B. McCabe; *dpp* loss of function alleles (*dpp^{hr4}* and *dpp^{hr56}*)^{24,25} were gifts from A. C. Spradling and T. E. Haerry, respectively; *oseg2^{Δ52}*, a null allele of *oseg2*, was a gift from T. Avidor-Reiss²⁶; *kfp10A^{Δ4}* null clones were generated by FLP/FRT-mediated removal of the rescue transgene in the background of *kfp10A* null mutation on the X chromosome: *kfp10A^{Δ4}/y*; *kfp10A P[acman]BAC* (CH322-03M08), *42DFRT/histoneGFP*, *42DFRT*; *hs-flp-MKRS/TM2*; *kfp10A^{Δ4}* (FBal0280190) was a gift from K. S. McKim²⁷. *P[acman]BAC* (CH322-03M08) transgenic flies were generated using strain BDSC24483 by PhiC31 integrase-mediated transgenesis (BestGene). Flies were heat-shocked at 37 °C for 1 h twice a day for 3 days and dissected after the indicated time.

RNAi screening of candidate genes for MT-nanotube morphology/function was performed by driving *UAS-RNAi* constructs under the control of *nos-gal4* (see below for validation method). Control crosses for RNAi screening were designed with matching *gal4* and *UAS* copy number using TriP background stocks (Bloomington Stock Center BDSC36304 or BDSC35787). Expression of *Dpp* under the *dpp-lexA* (LHG) driver or *tkv-lexA* driver (Bloomington Stock Center, BDSC54786) with *tub-gal80^{ts}* (denoted as *dpp-lexA^{ts}* or *tkv-lexA^{ts}*, respectively) was performed by culturing flies at 18 °C to avoid lethality during development and shifted to 29 °C upon eclosion for 24 h before analysis. For *shi^{ts}* expression, *nos-gal4>UAS-shi^{ts}* flies cultured at 18 °C were shifted to 29 °C upon eclosion for 24 h before analysis. The *dpp^{hr56}/CyO*; *nos-gal4*, *UAS-GFP-xtub* females were crossed with *dpp^{hr4}/SM6* males at permissive temperature (18 °C) and shifted to restrictive temperature (29 °C) upon eclosion for 24 h before analysis. We used *nos-gal4^{ts}* (see below for transgene construction) to achieve temporal control of *UAS-*tkv*-mCherry* to obtain similar expression levels to *dpp-lexA^{ts}>LexAop-dpp-GFP* for co-localization analysis and quantification of *Tkv-GFP* puncta. Temperature shift (29 °C) was performed upon eclosion for 72 h before analysis. Expression of *UAS-dome-EGFP²⁸* (a gift from S. Noselli) was performed at 18 °C with *nos-gal4* without *VP16* (see below for transgene construction). Other fly crosses were performed at 25 °C. Control experiments were conducted with matching temperature-shift schemes. For clonal expression of *Tkv-mCherry* and *GFP-αTub*, *hs-FLP*; *nos-FRT-stop-FRT-gal4*, *UAS-GFP²⁹* females were crossed with *UAS-*tkv*-mCherry*, *UAS-GFP-xtub* males and heat shocked at 37 °C for 20 min and observed 24 h after heat shock. A strong *tkv^{RNAi}* (TriP.HMS02185, Bloomington Stock Center BDSC40937) led to complete loss of germ cells, while a weak knockdown (TriP.GL01538, Bloomington Stock Center BDSC43194) partly maintained germ cells and was used for this study. Other stocks used in this study are listed in Extended Data Table 1.

Transgene construction. For constructions of *UAS-dlic-GFP*, *UAS-dlic-VN* and *UAS-kfp10A-VN*, open reading frames (ORF) were amplified from whole testis complementary DNA (cDNA) pool using polymerase chain reaction (PCR) with the following primer pairs. For *dlic*: 5'-ctagatctctcaaatggcgatgaacagtgaggacgca-3' and 5'-aactcgagacactcactctgcagatgtcaattttcacac-3'. These primers amplify BglII-Kozak sequence-*dlic*-XhoI fragment. For *kfp10A*: 5'-ctagatctctcaaatggcgatgattacgggtgg-3' and 5'-aactcgagacgcttgccattgcgcaattg-3'. These primers amplify BglII-Kozak sequence-*kfp10A*-XhoI fragment. (BglII and XhoI sites are indicated by underlines.)

Amplified fragments were sequenced for validation and subcloned into BglII/XhoI sites of *pUAST-EGFP-attB²⁹* or *pUAST-VenusN-attB* vector (containing the amino (N)-terminal half portion of Venus instead of GFP). *pUAST-VenusN-attB* vector was constructed as follows. The N-terminal half portion of Venus cDNA was amplified using primers, XhoI (underlined)-RSIAT (linker peptide, lower case)-Venus-F; 5'-AACTCGAGagatccattgcgacATGGTGAGCAAGGGCGA-3' and KpnI (underlined)-Venus-R; 5'-TCGGTACCTTAGGTGATATAGACGTTGTGGCTGATGTAGT-3' and subcloned into XhoI/KpnI sites of *pUAST-attB* vector³⁰. Transgenic flies were generated using strain BDSC24749 by PhiC31 integrase-mediated transgenesis (BestGene). *UAS-dlic-VN* and *UAS-kfp10A-VN* were used when GFP fluorescence was not necessary or undesirable. *UAS-kfp10A^{RNAi}* (double-stranded RNA HMS00920) target sequence is within the 5' untranslated region (UTR) of the gene and is not present in *UAS-kfp10A-VN* construct; thus this transgene was used to rescue RNAi-induced phenotypes (Extended Data Table 1).

To construct *nos-gal4* without *VP16*, Scer\GAL4 cDNA was amplified from the pG4PN-2 vector (a gift from C.-Y. Lee) using the following primers: NdeI (underlined)-Kozak-*gal4* ORF-F: 5'-aagcatatgttcaacatgaagctactgtctctatc-3' and EcoRI (underlined)-*gal4* ORF-R: 5'-tactcgaattctactcttttttgggtgg-3'. NheI-BamHI flanked the 3.13-kilobase fragment from pCSpnosFGVP (containing the *nanos* 5' UTR-ATG (NdeI start codon)); the *gal4-VP16-nanos* 3' UTR (a gift from E. R. Gavis) was subcloned into the NheI-BamHI site of the pUAST-attB vector³⁰. NdeI-EcoRI flanking *gal4-VP16* cassette was replaced by the NdeI-EcoRI-digested PCR fragment of Scer\GAL4 cDNA. Transgenes were introduced into the BDSC24482 strain using PhiC31 integrase-mediated transgenesis (BestGene). We used *nos-gal4* without *VP16* in combination with temperature-sensitive *gal80* (*tubulin-gal80^{ts}*), denoted as *nos-gal4^{ts}* to distinguish it from *nos-gal4-VP16* (ref. 31), which has been often referred to as *nos-gal4*.

Live imaging. For visualizing MT-nanotubes in unfixed samples, testes from newly eclosed flies (*nos-gal4>UAS-GFP-xtub*) were dissected in 1 ml of pre-warmed Schneider's *Drosophila* medium supplemented with 10% fetal bovine serum and glutamine-penicillin-streptomycin. Cytochalasin B (10 μM, Sigma-Aldrich) or colcemid (250 μM, Sigma-Aldrich) were added to the media and incubated at room temperature (22–23 °C) for 90 min. Hoechst 33342 (2 μg ml⁻¹) was added as necessary and incubated at room temperature for 30 min. Testes were washed twice with phosphate-buffered saline (PBS) before imaging. For MT-nanotube membrane visualization, testes were dissected in PBS, and FM4-64FX Lipophilic Styryl Dye (5 μg ml⁻¹, Molecular Probes) was added 1 min before analysis. Imaging was performed in the presence of dye within 15 min.

For time-lapse live imaging, testes were placed on a drop of medium on a microscope slide with coverslip spacers on both edges, and another coverslip was placed on top. Time-course images of the areas around hub were taken once every minute or every 5 min for 60 min using a Zeiss LSM700 confocal microscope with a ×40 oil immersion objective (numerical aperture = 1.4). Four-dimensional data sets (x, y, z, t) were processed using Image J³².

Immunofluorescent staining. Testes were dissected in PBS and fixed in 4% formaldehyde in PBS for 30 min. To preserve MT-nanotubes during fixation, taxol (1 μM) was added to 4% formaldehyde/PBS solution. For anti-α-tubulin staining, testes were fixed in 90% methanol, 3% formaldehyde for 10 min at –80 °C. Fixed testes were briefly rinsed three times and permeabilized in PBST (PBS + 0.3% Triton X-100) at room temperature for 1 h, followed by incubation with primary antibody in 3% bovine serum albumin (BSA) in PBST at 4 °C overnight. Samples were washed three times for 20 min in PBST, incubated with secondary antibody in 3% BSA in PBST at room temperature for 2–4 h and then washed for 60 min (three times 20 min) in PBST. Samples were then mounted using VECTASHIELD with 4',6'-diamidino-2-phenylindole (DAPI). The primary antibodies used were as follows: mouse anti-γ-tubulin (1:500; GTU-88, Sigma), rabbit anti-β-galactosidase (1:500, Abcam), rabbit anti-Klp10A (1:2,000, a gift from D. Sharp³³), rabbit anti-Ser⁴⁵³ and Ser⁴⁵⁵ phosphorylated Mad (1:1,000, a gift from E. Laufer), rat anti-Vg1a (1:20; developed by A. Spradling and D. Williams), mouse anti-Fascin III (7G10, 1:40, developed by C. Goodman), mouse anti-Armadillo (N2 7A1, 1:20, developed by E. Wieschaus) and mouse anti-α-tubulin 4.3 (1:50; developed by C. Walsh) were obtained from the Developmental Studies Hybridoma Bank. Guinea pig anti-STAT92E was generated using the synthetic peptide Ac-CSGTPHHAQESMQLGNGDFGMADFDTITNFENF-amide (Covance) and used at a dilution of 1:2,000. STAT92E antibody was validated by immunofluorescent staining of *nos-gal4^{ts}>stat92E^{RNAi}* (Bloomington Stock Center, BDSC35600 and BDSC33637, data not shown). Guinea pig anti-Klp10A was generated as described previously³³ (Covance) and used at a dilution of 1:2,000. AlexaFluor-conjugated secondary antibodies were used at a dilution of 1:400. Images were taken using a Zeiss LSM700 confocal microscope with a ×40 oil immersion objective (numerical aperture = 1.4), or a Leica TCS SP8 confocal microscope with a ×63 oil-immersion objective (numerical aperture = 1.4) and processed using Image J³² and Adobe Photoshop software. Three-dimensional rendering was performed by Imaris software.

Mosaic analysis and clonal knockdown. The *oseg2^{Δ52}* homozygous clones were generated by FLP/FRT-mediated mitotic recombination³⁴. Adult *hs-flp*, *tub-gal4*, *UAS-GFP*; *tub-gal80*, *2AFRT/oseg2^{Δ52}*, *2AFRT* males were heat-shocked at 37 °C for 1 h twice a day for 3 days; *hs-flp*, *tub-gal4*, *UAS-GFP*; *tub-gal80*, *2AFRT/2AFRT* flies were used as controls. Testes were dissected at indicated times after clone induction. The number of testes containing any GFP-positive *oseg2^{Δ52}* homozygous clones was determined. For RNAi clonal analysis, *hs-flp*; *nos-FRT-stop-FRT-gal4*, *UAS-GFP²⁹* with *UAS-che-13^{RNAi}*, *UAS-osm6^{RNAi}* or *UAS-kfp10A^{RNAi}* flies were heat-shocked at 37 °C for 30 min. Testes were dissected at indicated times after clone induction. The percentage of GFP-positive GSCs was determined. The means ± s.d. from two independent experiments were plotted to the graph.

Quantification of pMad and STAT92E intensities. For pMad quantification, integrated intensity within the GSC nuclear region was measured for anti-pMad staining and divided by the area. To normalize the staining condition, data were further normalized by the average intensities of pMad from four cyst cells in the same testes, and the ratios of relative intensities were calculated as each GSC per average cyst cell. For STAT92E quantification, intensity within the GSC nuclear region was measured for anti-STAT92E staining and divided by the area. Data were normalized by DAPI signal intensities. The means \pm s.d. were plotted to the graph for each genotype.

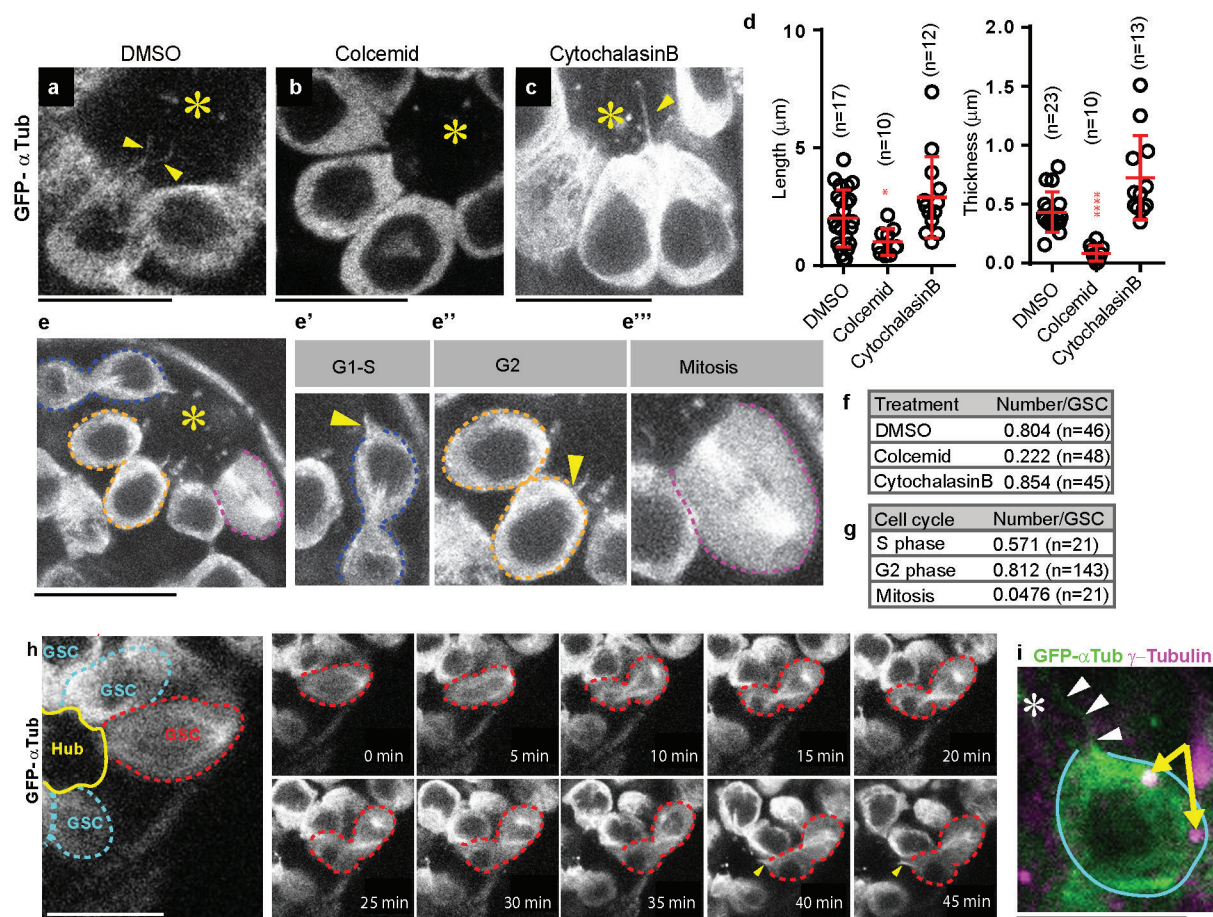
Quantitative reverse transcription PCR to validate RNAi-mediated knock-down of genes. Double driver females (*nos-gal4* and *c587-gal4*) were crossed with males of indicated RNAi lines. Testes from 50 male progenies, age 0–2 days, were collected and homogenized by pipetting in TRIzol Reagent (Invitrogen) and RNA was extracted following the manufacturer's instructions. One microgram of total RNA was reverse transcribed to cDNA using SuperScript III First-Strand Synthesis SuperMix (Invitrogen) with Oligo (dT)₂₀ Primer. Quantitative PCR was performed, in triplicate, using Cybergreen Applied Biosystems Gene Expression Master Mix on an CFX96 Real-Time PCR Detection System (Bio-Rad). Control primer for *xtub84B* (5'-TCAGACCTCGAAATCGTAGC-3'/5'-AGCAGTAGA GCTCCAGCAG-3') and experimental primer for *oseg1* (5'-TGATCATTCAG CACCTGATCTC-3'/5'-CGCCAGTCGATTCCGATAAA-3'), *oseg2* (5'-TCTG AACGAGCGAGGAAATG-3'/5'-CCACTGGTCATCCTGCTAATC-3'), *oseg3* (5'-ACTGGTTCTCGCAGGTAAAG-3'/5'-TAATGCCTCGCCAAGTGATAG-3'), *osm6* (5'-CTTCCATCCCAAGGAGTGATC-3'/5'-CTTCTCGTCACTGAA ATCGTAGT-3'), *che-13* (5'-GATGGAGCAGGAGCTGAAA-3'/5'-GGTCGG TGGTTTGGTTCT-3'), *tkv* (5'-GCCACGTCTCATCAACTCAA-3'/5'-CTTT GCACCAGCAATGGTAATC-3') were used. Relative quantification was performed using the comparative CT method (ABI manual).

Statistical analysis and graphing. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Statistical analysis and graphing were performed using Microsoft Excel 2010 or GraphPad prism 6 software. Data are shown as means \pm s.d. The *P* value (two-tailed Student's *t*-test) is provided for comparison with the control shown as **P* \leq 0.05, ***P* \leq 0.01, ****P* \leq 0.001; NS, non-significant (*P* > 0.05).

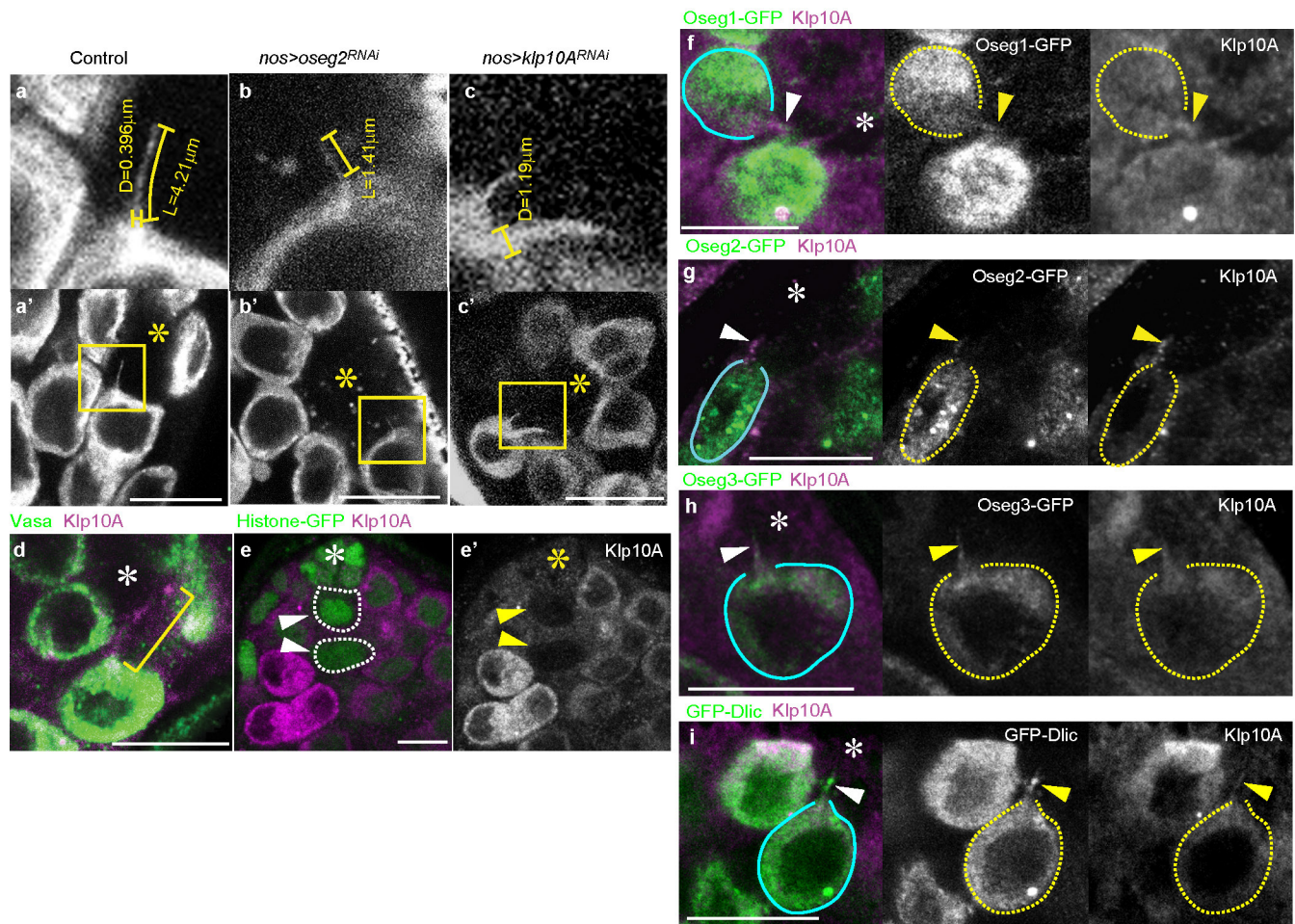
MT-nanotube orientation was measured as the angle between the two lines using imageJ³²: one formed by connecting germ cell centre to hub centre; the other formed by connecting the tip and the base of MT-nanotube. Each angle was plotted to Wind Rose graph by Origin 9.1 software (OriginLabs).

23. Yagi, R., Mayer, F. & Basler, K. Refined LexA transactivators and their use in combination with the *Drosophila* Gal4 system. *Proc. Natl Acad. Sci. USA* **107**, 16166–16171 (2010).
24. Wharton, K., Ray, R. P., Findley, S. D., Duncan, H. E. & Gelbart, W. M. Molecular lesions associated with alleles of decapentaplegic identify residues necessary for TGF- β /BMP cell signaling in *Drosophila melanogaster*. *Genetics* **142**, 493–505 (1996).
25. Xie, T. & Spradling, A. C. Decapentaplegic is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. *Cell* **94**, 251–260 (1998).
26. Avidor-Reiss, T. *et al.* Decoding cilia function. *Cell* **117**, 527–539 (2004).
27. Radford, S. J., Harrison, A. M. & McKim, K. S. Microtubule-depolymerizing kinesin KLP10A restricts the length of the acentrosomal meiotic spindle in *Drosophila* females. *Genetics* **192**, 431–440 (2012).
28. Ghiglione, C. The *Drosophila* cytokine receptor Domeless controls border cell migration and epithelial polarization during oogenesis. *Development* **129**, 5437–5447 (2002).
29. Salzmann, V., Inaba, M., Cheng, J. & Yamashita, Y. M. Lineage tracing quantification reveals symmetric stem cell division in *Drosophila* male germline stem cells. *Cell. Mol. Bioeng.* **6**, 441–448 (2013).
30. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
31. Van Doren, M., Williamson, A. L. & Lehmann, R. Regulation of zygotic gene expression in *Drosophila* primordial germ cells. *Curr. Biol.* **8**, 243–246 (1998).
32. ImageJ. (U.S. National Institutes of Health, 1997–2014).
33. Rogers, G. C. *et al.* Two mitotic kinesins cooperate to drive sister chromatid separation during anaphase. *Nature* **427**, 364–370 (2004).
34. Xu, T. & Rubin, G. M. Analysis of genetic mosaics in developing and adult *Drosophila* tissues. *Development* **117**, 1223–1237 (1993).
35. Grieder, N. C., de Cuevas, M. & Spradling, A. C. The fusome organizes the microtubule network during oocyte differentiation in *Drosophila*. *Development* **127**, 4253–4264 (2000).
36. Kuzhandaivel, A., Schultz, S. W., Alkhori, L. & Alenius, M. Cilia-mediated hedgehog signaling in *Drosophila*. *Cell Rep.* **7**, 672–680 (2014).



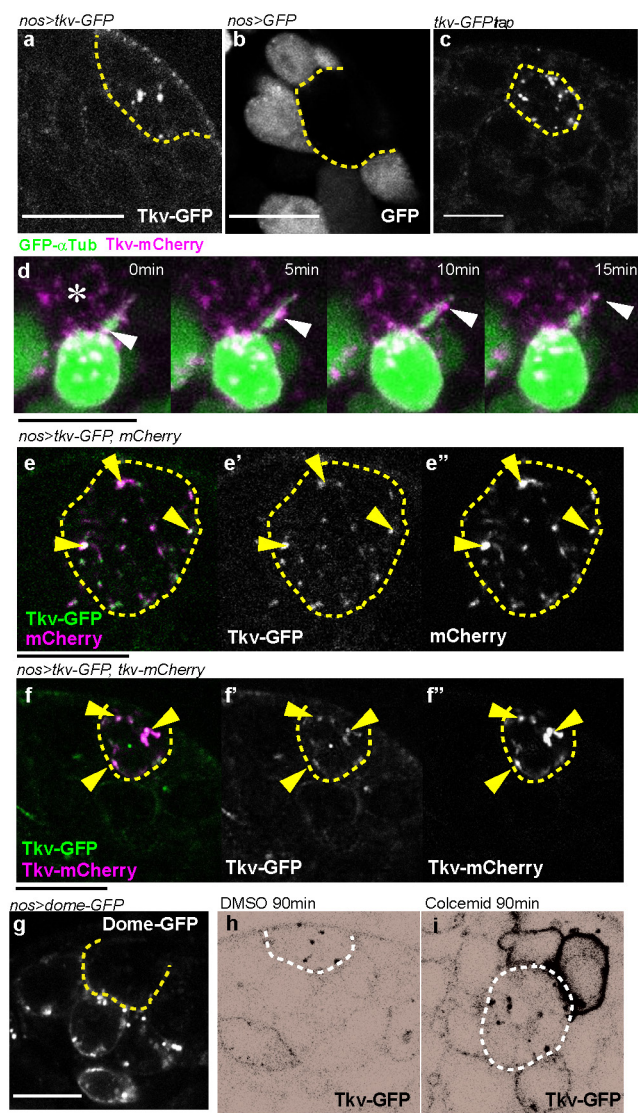
Extended Data Figure 1 | MT-nanotubes are MT-based structures that form in a cell-cycle-dependent manner. **a–c**, Representative images of MT-nanotubes visualized by GFP- α Tub (*nos-gal4>GFP- α tub*) after 90 min *ex vivo* treatment of mock (**a**, DMSO), colcemid (**b**) or cytochalasin B (**c**). **d**, Thickness and length of MT-nanotubes after mock (DMSO), colcemid or cytochalasin B treatment. Each scored value is plotted as an open circle. Red line indicates average value with standard deviation; *n* indicates the number of MT-nanotubes scored from more than three testes for each data point. **e**, Representative images of MT-nanotubes in each cell cycle stage visualized by GFP- α Tub. The three smaller panels to the right show magnified images of GSCs from the larger left-hand panel representing various stages of the cell cycle: left, G1–S phase (before the completion of the cytokinesis); middle, G2 phase; right, mitosis. **f**, **g**, Frequency of MT-nanotubes/GSC after mock

(DMSO), colcemid or cytochalasin B treatment (**f**) or during cell cycle (**g**); *n* indicates the number of GSCs scored from more than ten testes from three independent experiments for each data point. **h**, Frames from a time-lapse live imaging of a MT-nanotube visualized by GFP- α Tub. GSC in anaphase at 0 min is indicated by the red dotted circle, which undergoes cell division and grows MT-nanotubes (arrowheads) at 40 min (see Supplementary Video 1). MT-nanotubes typically formed during telophase to early S phase of the next cell cycle, within an hour after mitotic entry (95.2%, *n* = 21 GSCs) from three independent experiments. **i**, An example of a GSC that does not have the centrosome (arrows) at the base of the MT-nanotubes. MT-nanotubes are indicated by arrowheads. Centrosomes are indicated by arrows. Asterisk indicates hub. *P* values from *t*-tests are provided as **P* ≤ 0.05, ***P* ≤ 0.01, ****P* ≤ 0.001. Scale bar, 10 μm .

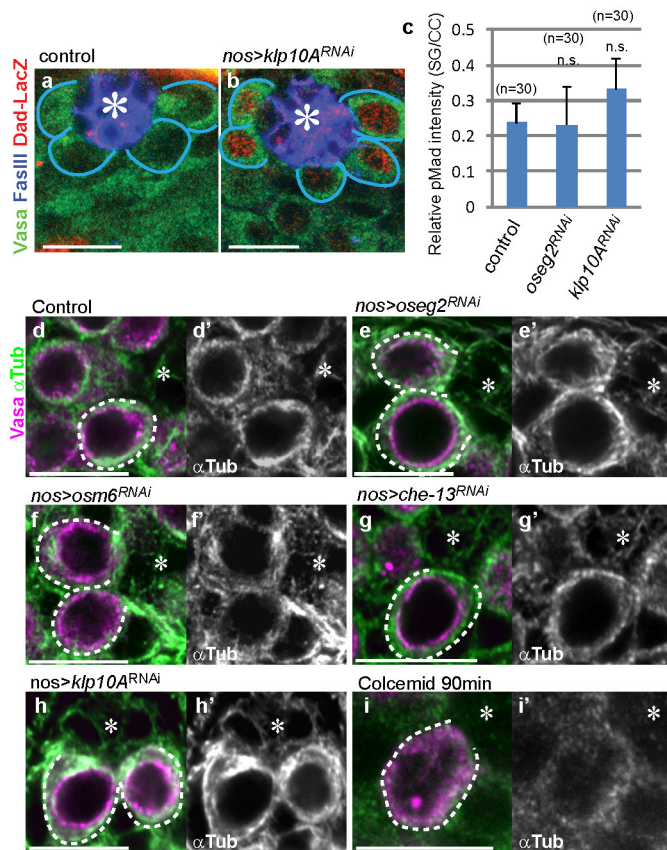


Extended Data Figure 2 | IFT proteins localize to MT-nanotubes. a–c, Examples of MT-nanotubes in wild type (a), *oseg2^{RNAi}* (b) and *klp10a^{RNAi}* (c) testes; *nos-gal4>GFP- α tub* was used. Upper panels are magnified views of squared areas in lower panels, showing examples of measuring length (L) and diameter (D, the base of the MT-nanotubes). d, An example of a MT-nanotube stained by anti-Klp10A antibody in WT testis. e, Validation of anti-Klp10A

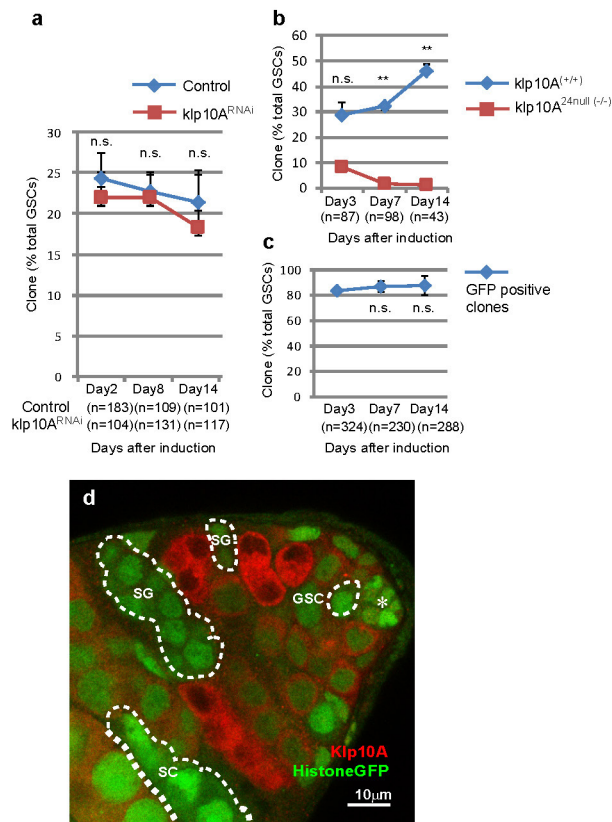
antibody, showing that *klp10a* mutant clones (arrowheads and dotted circles) have completely lost the staining 3 days after clone induction. f–i, Examples of testis apical tips expressing Oseg1-GFP (f), Oseg2-GFP (g), Oseg3-GFP (h), GFP-Dlic (i) driven by *nos-gal4*. Arrowheads indicate MT-nanotubes illuminated by anti-Klp10A staining. GSCs are indicated by blue lines or yellow dotted circles. Asterisk indicates hub. Scale bar, 10 μm.



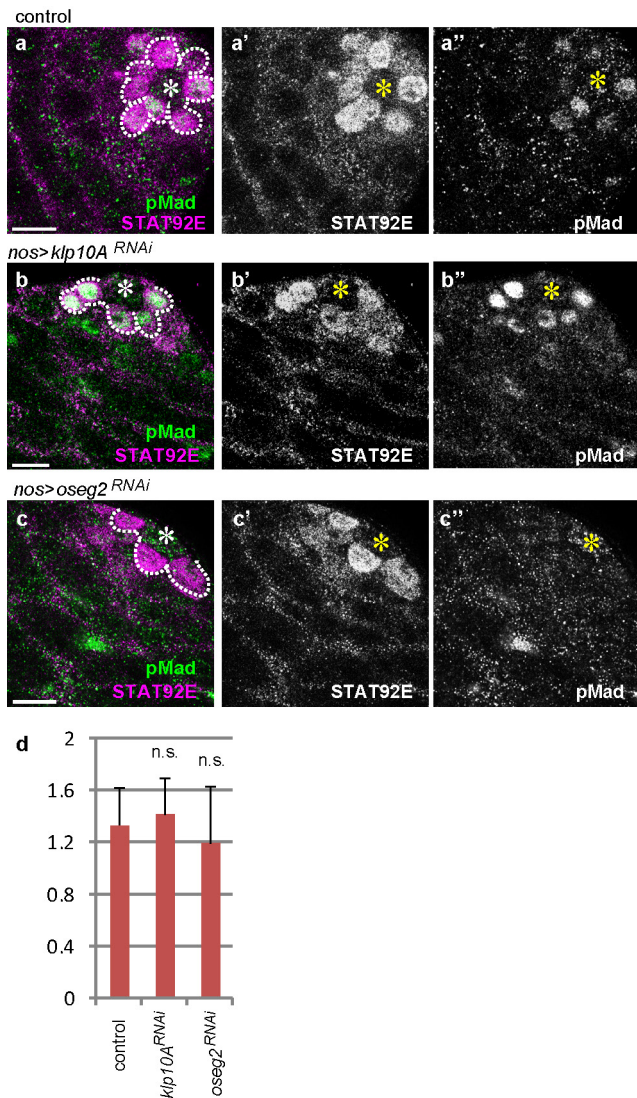
Extended Data Figure 3 | TkV-mCherry or mCherry co-localize with TkV-GFP in the hub. **a**, An apical tip of the testis expressing TkV-GFP in germ cells (*nos-gal4>tkv-GFP*). Broken lines indicate hub. **b**, An apical tip of the testis expressing GFP in germ cells (*nos-gal4>GFP*). **c**, Fully functional TkV-GFP protein trap shows punctate pattern within the hub area. **d**, Frames from a time-lapse live observation of TkV-mCherry puncta (arrowheads) moving along a MT-nanotube. Asterisk indicates hub. **e**, mCherry and TkV-GFP expressed in germ cells (*nos-gal4>UAS-tkv-GFP, UAS-mCherry*) co-localize in the hub (arrowheads). **f**, TkV-mCherry and TkV-GFP expressed together in germ cells (*nos-gal4>UAS-tkv-GFP, UAS-tkv-mCherry*) co-localize in the hub (arrowheads). **g**, An apical tip of the testis expressing Dome-GFP in germ cells (*nos-gal4>dome-GFP* raised at 18 °C to reduce the expression level). **h**, **i**, TkV-GFP localization in control (**h**, DMSO) or colcemid (**i**) treatment, revealing the localization of TkV-GFP to the GSC cortex upon perturbation of MT-nanotubes. Dotted hemi- or full circles indicate hub. Scale bar, 10 μ m.



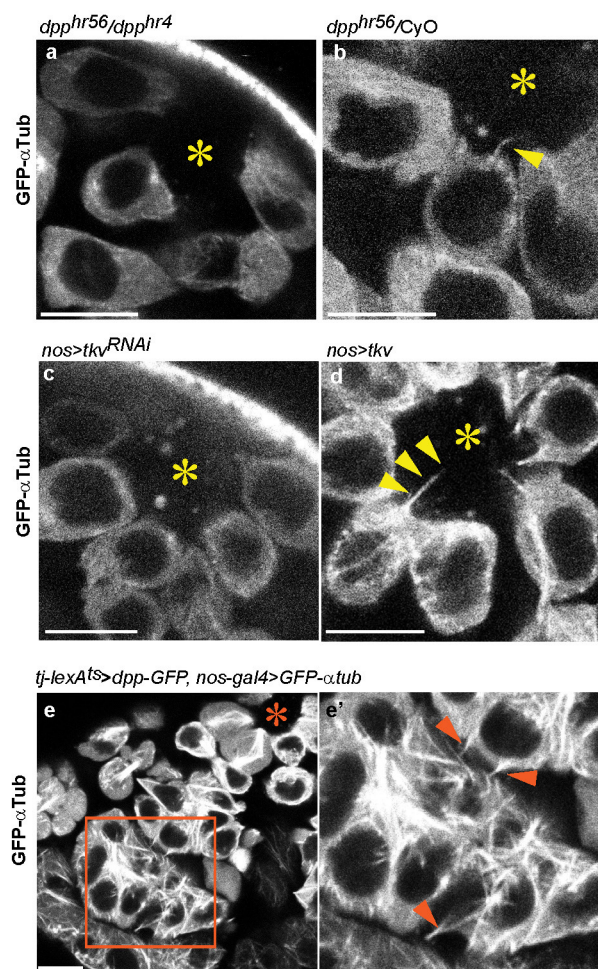
Extended Data Figure 4 | Effect of RNAi-mediated knockdown of IFT components on Dpp signalling and cytoplasmic microtubules. **a, b**, Dad-LacZ staining was undetectable in control GSCs (**a**) but was enhanced in *klp10A^{RNAi}* GSCs (**b**). **c**, Quantification of pMad intensity in the two- or four-cell spermatogonia (SG) of indicated genotypes. Graph shows average values \pm s.d.; $n = 30$ GSCs were scored from at least ten testes from at least two independent crosses for each data point. **d–i**, Cytoplasmic microtubule patterns stained with anti- α -tubulin antibody upon RNAi-mediated knockdown of indicated genes (**d–h**) or colcemid treatment for 90 min (**i**). In control as well as upon knockdown of IFT-B components, cytoplasmic MTs, visible as fibrous cytoplasmic patterns, were not visibly affected, whereas colcemid treatment disrupted cytoplasmic MTs; **h**, *klp10A* knockdown led to hyper stabilization of cytoplasmic MTs. Asterisk indicates hub. *P* values from *t*-tests are provided as NS, non-significant ($P > 0.05$). Scale bar, 10 μ m.



Extended Data Figure 5 | The *klp10A* mutant clones do not show a competitive advantage in GSC maintenance. **a**, Maintenance of *klp10A^{RNAi}* GSC clones. **b**, Maintenance of *klp10A^{24 null}* clones. The number of control GSC clones (+/+), determined by lack of GFP and the number of *klp10A^{24 null}* GSC clones (-/-), determined by anti-Klp10A staining) were scored. **c**, Maintenance of GFP-positive GSC clones from the cross of *42DFRT X histoneGFP*, *42DFRT*; *hs-flp-MKRS/TM2* as a control for *klp10A^{24 null}* clones in **b**. GFP-positive GSC clones did not decrease compared with day 3, excluding the possibility that *klp10A^{24 null}* GSC clones were lost because of unrelated mutation(s) on the *histoneGFP*, *42DFRT* chromosome. **a–c**, Indicated numbers of GSCs were scored for each data point from at least two independent crosses. **d**, A representative image of a testis with *klp10A^{24 null}* clones. The *klp10A^{24 null}* germ cells determined by anti-Klp10A staining are encircled by white dotted lines. Asterisk indicates hub. Scale bar, 10 μ m. Average values \pm s.d. are plotted in graphs. *P* values from *t*-tests are provided as * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$; NS, non-significant ($P > 0.05$).



Extended Data Figure 6 | STAT92E level is not affected by modulation of MT-nanotube formation. **a–c**, Double staining of STAT92E and pMad in control (**a**), *klp10A^{RNAi}* (**b**) and *oseg2^{RNAi}* (**c**) testes. Asterisk indicates hub. GSCs (and gonialblasts that are still connected to GSCs) are circled by the dotted line. **d**, Quantification of STAT92E intensity. GSCs ($n = 30$) from more than five testes from two independent crosses were scored for each data point. Average values \pm s.d. are shown. P values from t -test are provided as NS, non-significant ($P > 0.05$).



Extended Data Figure 7 | Dpp pathway is required for the MT-nanotube formation. **a, b**, Testes (**a**, dpp^{hr56}/dpp^{hr4}) or (**b**, dpp^{hr56}/CyO) expressing GFP- α Tub in germ cells ($nos-gal4 > GFP-\alpha tub$) at restrictive temperature. **c, d**, MT-nanotube formation upon knockdown (**c**) or overexpression (**d**) of Tkv visualized by GFP- α Tub. **e**, Ectopic MT-nanotube formation in spermatogonia upon expression of Dpp in somatic cyst cells. The right-hand panel is a magnified view of the squared region in the left panel. Arrowheads indicate ectopic MT-nanotubes. Asterisk indicates hub. Scale bar, 10 μ m.

Extended Data Table 1 | Effects of primary cilium or cytoneme genes on MT-nanotube formation

Transgenes	Sources, references	Thickness at the base (μm) [†] (number of nanotube scored)	Length (μm) [†] (number of nanotube scored)	MT-nanotube number /GSC (number of GSC scored from n>7 testes)	pMad level GSC/CC ^{†, ‡}	RNAi validation
<i>UAS-GFP-$\alpha\text{tub}^{\delta}$</i>	Bloomington stock center ³⁵	0.432 (51)	3.32 (82)	0.822 (584)	0.93 (34)	
<i>UAS-Dlic-VN</i>	This study	ND	ND	ND	1.01* (15)	
<i>UAS-Dlic^{RNAi} KK107892</i>	VDRC	0.412 (17)	1.91* (15)	0.519 (42)↓	0.534*** (15)	
<i>UAS-klp10A-VN</i>	This study	0.398 (39)	2.26* (21)	0.490 (79)↓	0.551*** (15)	
<i>UAS-klp10A^{RNAi} TRIP.HMS00920</i>	Bloomington stock center	0.827*** (105)	3.7 (65)	0.776 (58)	1.28** (40)	IHC
<i>UAS-klp10A^{RNAi}</i> with <i>UAS-klp10A-VN</i>	RNAi rescue	0.482 (40)	4.21* (40)	0.960 (75)	0.781 (15)	
<i>UAS-oseg1^{RNAi} KK101551</i>	VDRC	0.569 (16)	3.42 (15)	0.536 (41)	0.722 (15)	qPCR 0.57*
<i>UAS-oseg2^{RNAi} GD8122</i>	VDRC ³⁶	0.473 (25)	1.60*** (48)	0.194 (72)↓	0.448*** (44)	qPCR 0.68*
<i>UAS-oseg3^{RNAi} KK100864</i>	VDRC	0.552 (15)	2.51 (15)	0.663 (66)	0.707 (15)	qPCR 0.19***
<i>UAS-oseg1-GFP</i>	Avidor-Reiss, T ²⁶	ND	ND	ND	0.832 (15)	
<i>UAS-oseg2-GFP</i>	Avidor-Reiss, T ²⁶	0.482 (32)	3.99 (309)	0.727 (39)	1.14* (15)	
<i>UAS-oseg3-GFP</i>	Avidor-Reiss, T ²⁶	0.501 (15)	3.66 (15)	0.750 (37)	0.911 (15)	
<i>UAS-oseg4-GFP</i>	Avidor-Reiss, T ²⁶	ND	ND	ND	0.701 (15)	
<i>UAS-oseg5-GFP</i>	Avidor-Reiss, T ²⁶	0.255* (15)	2.62 (15)	0.333 (50) ↓	0.458*** (15)	
<i>UAS-osm6^{RNAi} GD24068</i>	VDRC	0.477 (15)	2.01* (17)	0.331 (51)↓	0.429*** (15)	qPCR 0.72**
<i>UAS-che-13^{RNAi} GD5096</i>	VDRC	0.494 (15)	2.16* (33)	0.306 (80)↓	0.449*** (30)	qPCR 0.20***
<i>dpp^{hr56}/dpp^{hr4} (29°C 24hr)</i>	Haerry, TE, Spradling, AC ^{24, 25}	0.173* (20)	4.30 (15)	0.0614 (244)↓	ND	
<i>dpp^{hr56}/dpp^{hr4} (18°C)</i>	Haerry, TE, Spradling, AC ^{24, 25}	0.412 (15)	3.02 (15)	0.545 (33)	ND	
<i>UAS-tkv-GFP</i>	Rodal, A ²²	0.378 (21)	4.89* (15)	0.784 (51)	1.25** (15)	
<i>UAS-tkv^{DN} (2d-GSK-3D2)</i>	Roy, S ¹⁸	0.779** (34)	3.67 (35)	0.649 (114)	0.679* (15)	
<i>UAS-tkv^{RNAi} TRIP.HMS02185</i>	Bloomington stock center	0.580 (15)	1.16** (15)	0.234 (90)↓	0.463** (15)	qPCR 0.76*
<i>UAS-caps^{DN}</i>	Roy, S ¹⁸	0.581 (15)	3.71 (15)	0.850 (60)	0.714 (15)	
<i>UAS-dia^{RNAi}</i>	Roy, S ¹⁸	0.407 (15)	3.34 (15)	0.667 (30)	0.991 (15)	
<i>UAS-shf^{ts} (29°C 24hr)</i>	Roy, S ¹⁸	0.536 (15)	3.22 (15)	0.632 (48)	0.894 (15)	
<i>UAS-shf^{ts} (18°C)</i>	Roy, S ¹⁸	0.481 (15)	3.78 (15)	0.701 (44)	0.811 (15)	

UAS-transgenes were driven by *nos-gal4* with *UAS-GFP- αtub* . ND, not determined. For all data points, data were obtained from at least two independent crosses.

[†] At least 15 randomly selected MT-nanotubes or at least 15 GSCs from more than seven testes were scored.

[‡] See Methods for quantification.

[§] Control.

^{||} See Methods for rescue experiment design.

P values from t-tests are provided as * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$; NS, non-significant ($P > 0.05$). Arrows are based on the significances ($P \leq 0.05$). References 35 and 36 are cited in the table.

Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria

Patrice Rassam^{1,2}, Nikki A. Copeland^{2†}, Oliver Birkholz³, Csaba Tóth^{2†}, Matthieu Chavent¹, Anna L. Duncan¹, Stephen J. Cross^{2†}, Nicholas G. Housden¹, Renata Kaminska¹, Urban Seger^{2†}, Diana M. Quinn², Tamsin J. Garrod^{2†}, Mark S. P. Sansom¹, Jacob Piehler³, Christoph G. Baumann² & Colin Kleanthous¹

Gram-negative bacteria inhabit a broad range of ecological niches. For *Escherichia coli*, this includes river water as well as humans and animals, where it can be both a commensal and a pathogen^{1–3}. Intricate regulatory mechanisms ensure that bacteria have the right complement of β -barrel outer membrane proteins (OMPs) to enable adaptation to a particular habitat^{4,5}. Yet no mechanism is known for replacing OMPs in the outer membrane, an issue that is further confounded by the lack of an energy source and the high stability⁶ and abundance of OMPs⁵. Here we uncover the process underpinning OMP turnover in *E. coli* and show it to be passive and binary in nature, in which old OMPs are displaced to the poles of growing cells as new OMPs take their place. Using fluorescent colicins as OMP-specific probes, in combination with ensemble and single-molecule fluorescence microscopy *in vivo* and *in vitro*, as well as molecular dynamics simulations, we established the mechanism for binary OMP partitioning. OMPs clustered to form $\sim 0.5\text{-}\mu\text{m}$ diameter islands, where their diffusion is restricted by promiscuous interactions with other OMPs. OMP islands were distributed throughout the cell and contained the Bam complex, which catalyses the insertion of OMPs in the outer membrane^{7,8}. However, OMP biogenesis occurred as a gradient that was highest at mid-cell but largely absent at cell poles. The cumulative effect is to push old OMP islands towards the poles of growing cells, leading to a binary distribution when cells divide. Hence, the outer membrane of a Gram-negative bacterium is a spatially and temporally organized structure, and this organization lies at the heart of how OMPs are turned over in the membrane.

We developed a strategy for following the localization and turnover of OMPs using colicins^{9,10}, which circumvents the need for fluorescent protein fusions. Protein fusions are the method-of-choice for investigating the localization of cytoplasmic, inner membrane and periplasmic proteins¹¹, but these generally inhibit OMP maturation. Colicins are species-specific bacteriocins that bind OMP receptors before translocating a cytotoxic domain into the cell⁹. ColE9 and ColIa were used here as specific, high-affinity ($\sim\text{nM}$), non-covalent labels for the vitamin B₁₂ transporter BtuB and the iron siderophore transporter Cir, respectively^{12–14} (Fig. 1a). The colicins were engineered with disulfide bonds to block their import into bacterial cells¹² and covalently modified with organic fluorophores (Alexa Fluor 488 (AF488) or tetramethyl rhodamine (TMR)). We first compared the distribution and turnover of ColE9^{TMR}-labelled BtuB in the outer membrane with an inner membrane protein (IMP), green fluorescent protein (GFP)-labelled TatA, by confocal fluorescence microscopy following simultaneous, pulsed production of both proteins. These experiments were carried out over several rounds of cell division using *E. coli* JM83 cells expressing *btuB* and *tatA-GFP*¹⁵ genes from arabinose-inducible promoters, in this instance using a variant of ColE9 in which

the amino-terminal 83 amino acids are deleted ($\Delta^{1–83}$ ColE9^{TMR}) to label BtuB. Unlike TatA–GFP, which showed a purely analogue diminution of fluorescence in the inner membrane as cells divided, BtuB–ColE9^{TMR} distribution was binary in nature (Fig. 1b, f and Supplementary Video 1). The OMP migrated towards the old poles during the first cell division and thereafter was sequestered predominantly in two cells, which we designated repository cells. Repository cells retained the bulk of the transiently produced OMP (BtuB), leaving daughter cells to acquire new (unlabelled) OMPs. To ensure transient expression had not influenced OMP behaviour, we repeated these experiments but this time observed endogenous BtuB and Cir migration in JM83 cells, following labelling with the respective colicin (Supplementary Video 2). Starting from a single cell, the two OMPs behaved identically. Approximately half of each OMP moved towards the poles as the cell grew and divided and were then retained in two repository cells, while new daughter cells received little or no old colicin-labelled OMP.

We used a combination of laser scanning confocal fluorescence recovery after photobleaching (FRAP) and total internal reflection fluorescence microscopy (TIRFM) to determine the mechanistic basis of binary OMP partitioning in *E. coli*. Fluorescence recovery was rapid for TatA–GFP ($<3\text{ s}$ at 37°C), which is typical of IMPs¹⁵, whereas no recovery was observed for colicin-labelled BtuB or Cir even after 3 min (Extended Data Fig. 1), similar to the OMP OmpA¹⁶. From single-molecule tracking (SMT)-TIRFM, we established that BtuB and Cir have planar diffusion coefficients ($D \sim 0.013$ and $0.019\text{ }\mu\text{m}^2\text{ s}^{-1}$, respectively) that are similar to those reported previously for IMPs and OMPs *in vivo* (Extended Data Fig. 2a, b, Extended Data Table 1 and Supplementary Videos 3 and 4), demonstrating that colicin-labelled OMPs diffuse in the membrane. However, mean square displacement (MSD) plots indicated that this diffusion was restricted to regions of $\sim 0.5\text{ }\mu\text{m}$ diameter. Analysis of various *E. coli* mutants indicated that cell envelope structures such as cross-bridges to the peptidoglycan were not responsible for the restricted diffusion of OMPs (Extended Data Fig. 3 and Extended Data Table 1). However, Monte Carlo simulations of BtuB diffusion were consistent with its confinement being due to an increase in mass over time, such as through association with other OMPs (Extended Data Fig. 2c), a point we return to later.

We next investigated the distribution and co-localization of colicin-labelled BtuB and Cir in *E. coli* JM83 cells using TIRFM. BtuB–ColE9^{AF488} and Cir–ColIa^{TMR} clustered together in islands that were distributed throughout the cell and moved to the poles as cells divided (Fig. 1c, d). Similar punctate appearance of trimeric porins in the outer membrane of *E. coli* has previously been reported^{17,18}. OMP islands contained multiple copies of BtuB and Cir (mean ~ 7 ; Extended Data Fig. 4a) and their mean size was $\sim 0.5\text{ }\mu\text{m}$ (Fig. 1e), equivalent to the

¹Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. ²Department of Biology, University of York, York YO10 5DD, UK. ³Department of Biology, University of Osnabrück, Barbarastraße 11, 49076 Osnabrück, Germany. [†]Present addresses: Biomedical and Life Sciences, Faculty of Health and Medicine, Lancaster University, Lancaster LA1 4YQ, UK (N.A.C.); Biomedica Slovakia s.r.o., Drobňého 27, 841 01 Bratislava, Slovakia (C.T.); School of Biochemistry, Medical Sciences Building, University Walk, Clifton, Bristol BS8 1TD, UK (S.J.C.); medi | Zentrum für medizinische Bildung, Max-Daetwyler-Platz 2, 3014 Bern, Switzerland (U.S.); Basil Hetzel Institute, Adelaide University, Woodville Road, Adelaide, South Australia 5011, Australia (T.J.G.).

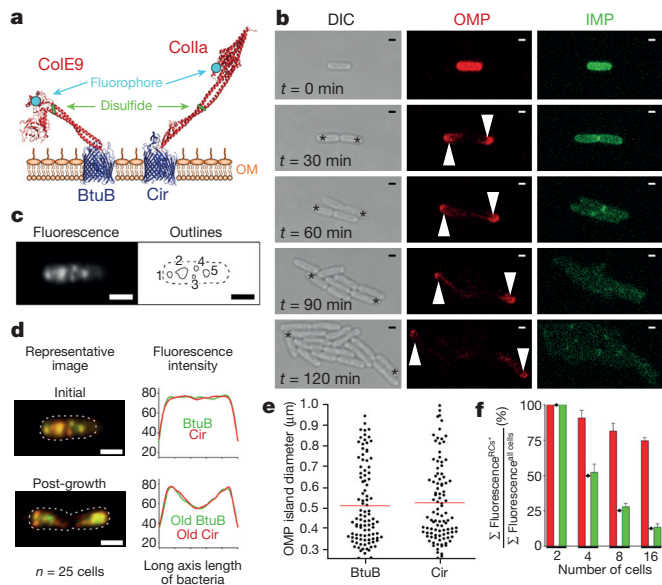


Figure 1 | Binary partitioning of OMPs in the *E. coli* outer membrane.

a, Structures of colicins bound to their OMP receptors (Protein Data Bank (PDB) accessions 1UJW, 1CJH and 1FSJ for ColE9–BtuB model, and 2HDI and 1CII for ColIa–Cir) highlighting the positions of inactivating disulfide bonds and fluorophore labels. OM, outer membrane. **b**, Distribution of fluorescently labelled OMP (BtuB- Δ^{1-83} ColE9^{TMR}) and IMP (TatA–GFP) in *E. coli* JM83 cells following transient induction with arabinose and imaging by confocal microscopy (see Methods). In contrast to TatA–GFP, which distributed in a purely analogue fashion (see **f**), BtuB- Δ^{1-83} ColE9^{TMR} segregation was binary. After four divisions two OMP repository cells (white arrowheads or asterisks) retained most of the original OMP. DIC, differential interference contrast. **c**, Left, TIRFM image (sum of 100 frames) of BtuB- and Cir-containing OMP islands in *E. coli* JM83 cells labelled with ColE9^{AF488} and ColIa^{TMR}, respectively. Right, automated ImageJ delineation of OMP islands. **d**, Top, false colour image of endogenous, colicin-labelled BtuB (green) and Cir (red) showing co-localization in OMP islands (same cell as in **c**). Bottom, as above but cells were grown for 1 h before imaging of cells undergoing division. Of the 130 OMP islands analysed before and after growth ($n = 20$ cells per experiment and in duplicate), $40 \pm 9\%$ and $39 \pm 7\%$ (mean \pm s.e.m.), respectively, contained both BtuB and Cir. All scale bars, 1 μ m. **e**, Size distribution of BtuB- and Cir-containing OMP islands from **c** and **d**. Red line denotes the mean. **f**, Distribution of fluorescence, shown as a ratio of the two repository cells relative to total fluorescence, at each cell division for TatA–GFP (green bars) and BtuB- Δ^{1-83} ColE9^{TMR} (red bars) from five confocal microscopy data sets. Data are mean and s.e.m. Diamond symbols denote expected ratios for an analogue mechanism.

confinement diameter observed in SMT-TIRFM experiments. Linking these two observations we suggest restricted diffusion of OMPs in the outer membrane reflects confinement within OMP islands. Taking the cross-sectional diameters of the BtuB and Cir β -barrels (~ 50 Å) as the average for an OMP, these data further suggest that a typical island comprises many hundreds of OMPs. Hence, OMP islands contain other OMPs not just BtuB and Cir.

Focusing on BtuB, we compared the behaviour of new and old OMPs in the outer membrane. ColE9^{TMR} (red) and ColE9^{AF488} (green) were used to label old and new BtuB, respectively, where the addition of labels was staggered by a growth phase (Fig. 2 and Supplementary Video 5). Temporal separation of the labels yielded poor or no co-localization in both confocal and TIRFM experiments. Old BtuB-containing islands accumulated at the poles of dividing cells while new OMP islands took their place. The new OMPs moved to the poles in the next division as they became old OMPs (Fig. 2a). Notably, these data highlighted a gradient of OMP biogenesis that was greatest at mid-cell and diminished towards the poles (Fig. 2b). The movement of old OMP islands to the poles is the result of this asymmetric biosynthesis. The TIRFM data also demonstrated the stochastic nature of

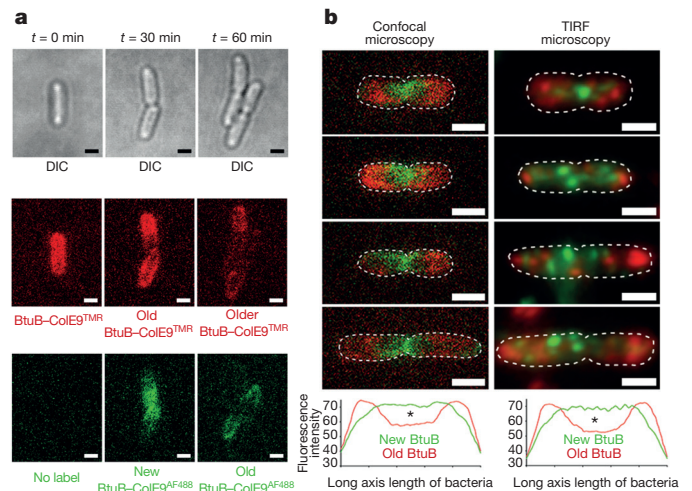


Figure 2 | Biogenesis of new OMP islands in the central regions of dividing *E. coli* cells forces old OMP islands towards the poles. **a**, Sequential insertion of OMPs in the outer membrane followed by two-colour confocal fluorescence microscopy. Growing *E. coli* JM83 cells were labelled initially with ColE9^{TMR} (red) ($t = 0$ min) and then with ColE9^{AF488} (green) ($t = 30$ min), followed by a final growth phase ($t = 60$ min). Representative data shown from duplicate experiments. **b**, **c**, Confocal and TIRF microscopy images, respectively, of individual dividing *E. coli* JM83 cells from the 30-min time point (**a**). The mean fluorescence distributions shown below the panels are for 20 cells. $*P < 0.1$, Mann–Whitney test. Scale bars, 1 μ m. Of the 165 OMP islands analysed by TIRFM at the 30-min time point ($n = 20$ cells per experiment and in duplicate) only $6.4 \pm 6.2\%$ showed overlap of old/new BtuB fluorescence, emphasizing that old OMP islands do not contribute to new OMP synthesis.

OMP biogenesis, with cell-to-cell variation in both number and distribution of new OMP islands.

OMP islands are deposited in the outer membrane of Gram-negative bacteria by the Bam machinery although where or how this occurs is not known⁷. We found that components of the Bam complex (BamA and BamC), which were labelled with fluorescently labelled antibodies^{19,20}, showed significant co-localization with BtuB- and Cir-containing OMP islands, and that this co-localization persisted as islands migrated towards the poles (Fig. 3a, b and Extended Data Fig. 4b, c). Importantly, new Bam-containing islands appeared in non-polar regions of dividing cells consistent with these regions being the major sites of OMP biogenesis (Fig. 2b).

The co-localization of different OMPs (BtuB, Cir and BamA) within islands and their restricted diffusion suggested promiscuous protein–protein interactions (PPIs) might be the underlying cause. To test this hypothesis, we reconstituted BtuB in a polymer-supported membrane (PSM)²¹ prepared from an *E. coli* lipid extract and followed the lateral mobility of single molecules after labelling with ColE9^{TMR} (see Methods). At low BtuB densities (~ 50 BtuB per μ m²), similar to that estimated for the outer membrane²², most of the BtuB–ColE9^{TMR} complexes exhibited Brownian diffusion with trajectories that extended beyond those observed *in vivo* (Fig. 4a, b, panel 1 and Supplementary Video 6) and with faster diffusion coefficients ($D \sim 0.18 \mu$ m² s^{−1}). Over time, however, BtuB molecules began to diffuse more slowly suggesting self-association (Fig. 4b, panel 2 and Extended Data Fig. 5a). This interpretation was confirmed by experiments in which BtuB was added at a 1,000-fold higher concentration (Fig. 4b, panel 3). Addition of the trimeric porin OmpF at concentrations found in the outer membrane of *E. coli* (1,000-fold above that of BtuB) resulted in most of the BtuB–ColE9^{TMR} complexes exhibiting restricted diffusion similar to that seen *in vivo* ($D \sim 0.02 \mu$ m² s^{−1} and confinement diameter ~ 0.4 μ m, respectively; Fig. 4b, panel 4 and Extended Data Fig. 2d). Addition of maltose binding protein, which had been engineered with a single transmembrane helix (TM-MBP), to the PSM (Fig. 4b, panel 5; Extended Data Fig. 5b, d) or lipopolysaccharide (LPS) (Extended

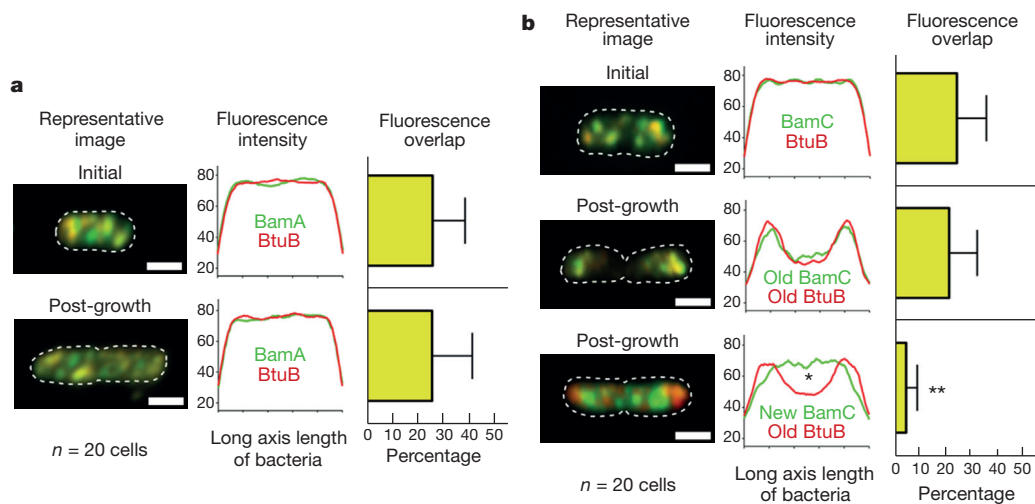


Figure 3 | The Bam biogenesis machinery is located within OMP islands that emerge primarily in non-polar regions of the cell and migrate to the poles. **a**, TIRFM images (sum of 100 frames) of *E. coli* JWD3 cells producing haemagglutinin (HA)-tagged BamA detected by Alexa⁴⁸⁸-labelled anti-HA antibody¹⁹ (see Methods). ColE9^{TMR} was used to detect endogenous BtuB before and after a 1-h period of growth. Of the 115 OMP islands analysed, $33 \pm 9\%$ (mean \pm s.e.m.) contained both BamA and BtuB. **b**, TIRFM images

(sum of 100 frames) for *E. coli* JM83 cells stained with Alexa⁴⁸⁸-labelled anti-BamC antibody and ColE9^{TMR} (BtuB). Temporal separation of labels by a 1-h period of growth showed old BtuB-containing islands localized primarily at the poles (red label), whereas new BamC-containing islands localized in non-polar regions (green label). The mean fluorescence distributions and colocalization histograms (error reported as s.e.m.) shown in the panels are from 20 cells in each case. * $P < 0.1$ (Student's *t*-test); ** $P < 0.001$ (Mann-Whitney).

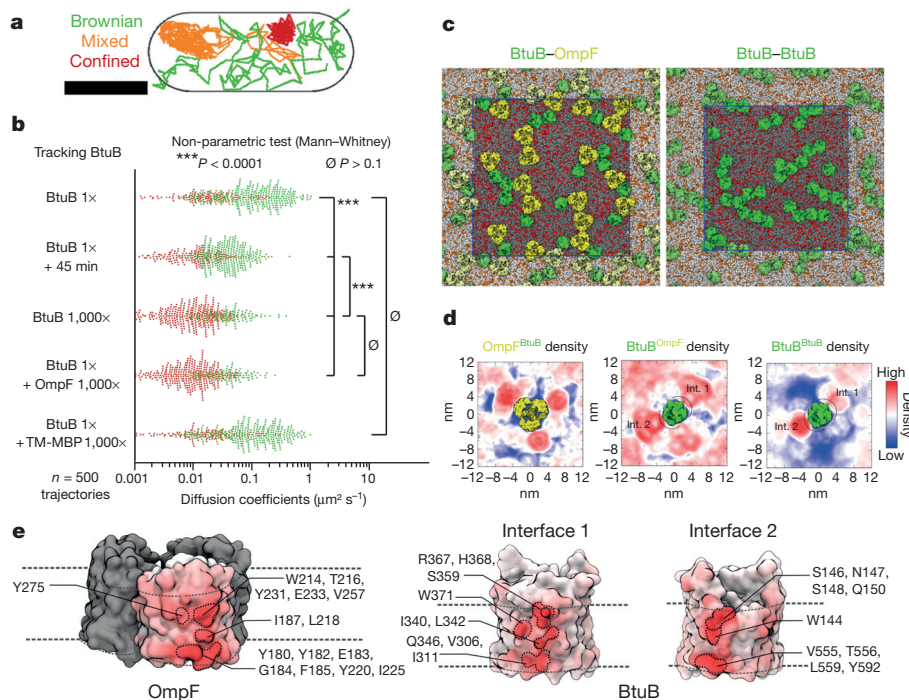


Figure 4 | OMPs engage in promiscuous protein-protein interactions. **a**, Representative single molecule trajectories (from duplicate experiments) of ColE9^{TMR}-labelled BtuB reconstituted into a PSM overlaid onto a scaled outline of an *E. coli* cell (see Supplementary Video 6). Molecules exhibited Brownian diffusion, typically extending beyond the boundaries of an *E. coli* cell (green trace), mixed (orange trace) or confined diffusion (red trace). Scale bar, 1 μm . See Methods. **b**, Initially (at 5-min post-deposition of the PSM), and at concentrations found in the outer membrane of *E. coli* (denoted by 1 \times), BtuB molecules displayed predominantly Brownian diffusion (panel 1), but with faster diffusion coefficients than observed *in vivo*. Over time, however, BtuB exhibited slower diffusion suggesting self-association (panel 2), an interpretation that was confirmed by raising the BtuB concentration significantly (panel 3) which resulted in most of the molecules exhibiting restricted diffusion. The same effect could be elicited using an unrelated OMP, OmpF (panel 4), but not maltose binding protein (TM-MBP) tethered to the

membrane through a single transmembrane helix (panel 5). **c**, Snapshots of a crowded bilayer ($\sim 30\%$ and $\sim 20\%$ protein fractional area, for BtuB-OmpF and BtuB-BtuB simulations, respectively) from molecular dynamics simulations after 10 μs (see also Extended Data Figs 6 and 7 and Extended Data Table 2). OmpF is shown in yellow, BtuB in green. The blue inner square represents the simulation box ($60 \times 60 \text{ nm}^2$) while the outer region represents the periodic replicates. Lipids in the unit cell were present in a ratio of 3:1 phosphatidylethanolamine (PE; grey) and phosphatidylglycerol (PG; red). **d**, Density plots of the frequency of occurrence of BtuB around OmpF (left) and of OmpF around BtuB (middle) observed in the BtuB-OmpF simulations, and of BtuB around BtuB (right) in the BtuB-BtuB simulations (see Extended Data Fig. 7). **e**, High frequency interaction sites at the interface between BtuB and OmpF. Dashed lines represent the position of the bilayer. The OmpF-BtuB interactions were predominantly mediated by hydrophobic and aromatic residues and, to a lesser extent, polar residues.

Data Fig. 5b, c) did not induce BtuB clustering, indicating that this effect was not simply due to crowding or LPS association, but involved interactions between OMPs. Recent atomic force microscopy data and molecular dynamics simulations indicate OMPs can self-associate in membranes^{23–25}. Using coarse-grained molecular dynamics simulations of BtuB and OmpF/BtuB mixtures at high local concentration, we further showed that OMPs have a propensity to engage in both homologous and heterologous associations that slow their diffusion (Fig. 4c–e, Extended Data Figs 6 and 7 and Extended Data Table 2). Promiscuous OMP–OMP interactions were largely, but not exclusively, mediated by aromatic residues displayed from the surfaces of their β -barrels. In the case of BtuB, the same set of residues mediated both homologous and heterologous interactions (Fig. 4e). We propose that OMP islands are stabilized *in vivo* by such promiscuous PPIs.

In conclusion, we have shown that biogenesis and turnover of OMPs in the Gram-negative bacterium *E. coli* are inextricably linked processes (see model in Extended Data Fig. 8). The resulting binary partitioning of OMPs ensures they are turned over rapidly, but also means old OMPs persist in repository cells, which implies bacterial populations have ‘memory’ of past OMP biosynthesis. The accumulation of old OMPs in repository cells could also be a factor in cellular ageing as well as providing a route for the presentation of proteins at cell poles^{26,27}. The co-localization of the Bam biogenesis machinery with the OMPs it has deposited in the membrane points to a high level of cellular organization and coordination in the outer membrane that has hitherto been unsuspected. Bam substrates include LptD (ref. 28), which inserts LPS into the outer leaflet of the outer membrane, and the β -barrels of autotransporters²⁹, which are crucial for biofilm formation and adhesion to host epithelia during bacterial pathogenesis. Moreover, since Bam is also required for the surface exposure of the outer membrane lipoprotein RcsF through the pores of OMPs³⁰, OMP and lipoprotein biogenesis may intersect within OMP islands.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 September 2014; accepted 8 April 2015.

Published online 10 June 2015.

- van Elsas, J. D., Semenov, A. V., Costa, R. & Trevors, J. T. Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *ISME J.* **5**, 173–183 (2011).
- Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal *Escherichia coli*. *Nature Rev. Microbiol.* **8**, 207–217 (2010).
- Morabito, S. (ed.) *Pathogenic Escherichia coli: Molecular & cellular microbiology* (Caister Academic Press, 2014).
- Vogel, J. & Papenfort, K. Small non-coding RNAs and the bacterial outer membrane. *Curr. Opin. Microbiol.* **9**, 605–611 (2006).
- Nikaido, H. Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* **67**, 593–656 (2003).
- Moon, C. P., Zaccari, N. R., Fleming, P. J., Gessmann, D. & Fleming, K. G. Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. *Proc. Natl Acad. Sci. USA* **110**, 4285–4290 (2013).
- Hagan, C. L., Silhavy, T. J. & Kahne, D. β -Barrel membrane protein assembly by the Bam complex. *Annu. Rev. Biochem.* **80**, 189–210 (2011).
- Noinaj, N. *et al.* Structural insight into the biogenesis of β -barrel membrane proteins. *Nature* **501**, 385–390 (2013).
- Kleanthous, C. Swimming against the tide: progress and challenges in our understanding of colicin translocation. *Nature Rev. Microbiol.* **8**, 843–848 (2010).
- Housden, N. G. *et al.* Intrinsically disordered protein threads through the bacterial outer membrane porin OmpF. *Science* **340**, 1570–1574 (2013).
- Shapiro, L., McAdams, H. H. & Losick, R. Why and how bacteria localize proteins. *Science* **326**, 1225–1228 (2009).
- Housden, N. G., Loftus, S. R., Moore, G. R., James, R. & Kleanthous, C. Cell entry mechanism of enzymatic bacterial colicins: porin recruitment and the thermodynamics of receptor binding. *Proc. Natl Acad. Sci. USA* **102**, 13849–13854 (2005).
- Kurusu, G. *et al.* The structure of BtuB with bound colicin E3 R-domain implies a translocon. *Nature Struct. Biol.* **10**, 948–954 (2003).
- Buchanan, S. K. *et al.* Structure of colicin I receptor bound to the R-domain of colicin Ia: implications for protein import. *EMBO J.* **26**, 2594–2604 (2007).
- Mullineaux, C. W., Nenniger, A., Ray, N. & Robinson, C. Diffusion of green fluorescent protein in three cell environments in *Escherichia coli*. *J. Bacteriol.* **188**, 3442–3448 (2006).
- Verhoeven, G. S., Dogterom, M. & den Blaauwen, T. Absence of long-range diffusion of OmpA in *E. coli* is not caused by its peptidoglycan binding domain. *BMC Microbiol.* **13**, 66 (2013).
- Smit, J. & Nikaido, H. Outer membrane of Gram-negative bacteria. XVIII. Electron microscopic studies on porin insertion sites and growth of cell surface of *Salmonella typhimurium*. *J. Bacteriol.* **135**, 687–702 (1978).
- Ursell, T. S., Trepagnier, E. H., Huang, K. C. & Theriot, J. A. Analysis of surface protein expression reveals the growth pattern of the Gram-negative outer membrane. *PLOS Comput. Biol.* **8**, e1002680 (2012).
- Browning, D. F. *et al.* Mutational and topological analysis of the *Escherichia coli* BamA protein. *PLoS ONE* **8**, e84512 (2013).
- Webb, C. T. *et al.* Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC. *J. Mol. Biol.* **422**, 545–555 (2012).
- Roder, F. *et al.* Reconstitution of membrane proteins into polymer-supported membranes for probing diffusion and interactions by single molecule techniques. *Anal. Chem.* **83**, 6792–6799 (2011).
- White, J. C., DiGirolamo, P. M., Fu, M. L., Preston, Y. A. & Bradbeer, C. Transport of vitamin B₁₂ in *Escherichia coli*. Location and properties of the initial B₁₂-binding site. *J. Biol. Chem.* **248**, 3978–3986 (1973).
- Casuso, I. *et al.* Characterization of the motion of membrane proteins using high-speed atomic force microscopy. *Nature Nanotechnol.* **7**, 525–529 (2012).
- Goose, J. E. & Sansom, M. S. Reduced lateral mobility of lipids and proteins in crowded membranes. *PLOS Comput. Biol.* **9**, e1003033 (2013).
- Jaroslowski, S., Duquesne, K., Sturgis, J. N. & Scheuring, S. High-resolution architecture of the outer membrane of the Gram-negative bacteria *Roseobacter denitrificans*. *Mol. Microbiol.* **74**, 1211–1222 (2009).
- Stewart, E. J., Madden, R., Paul, G. & Taddei, F. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* **3**, e45 (2005).
- Laloux, G. & Jacobs-Wagner, C. How do bacteria localize proteins to the cell pole? *J. Cell Sci.* **127**, 11–19 (2014).
- Qiao, S., Luo, Q., Zhao, Y., Zhang, X. C. & Huang, Y. Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. *Nature* **511**, 108–111 (2014).
- Ieva, R., Tian, P., Peterson, J. H. & Bernstein, H. D. Sequential and spatially restricted interactions of assembly factors with an autotransporter β domain. *Proc. Natl Acad. Sci. USA* **108**, E383–E391 (2011).
- Konovalova, A., Perlman, D. H., Cowles, C. E. & Silhavy, T. J. Transmembrane domain of surface-exposed outer membrane lipoprotein RcsF is threaded through the lumen of β -barrel proteins. *Proc. Natl Acad. Sci. USA* **111**, E4350–E4358 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the University of Oxford Micron facility and the University of York Biology Technology Facility for access to microscopy facilities. The authors also thank S. Buchanan and K. Jakes for anti-BamC antibody and Colla plasmids, I. Henderson for HA-tagged BamA constructs, C. Robinson for Tata-GFP construct, K. Heurlier and J. Pullen for PCR screening of deletion strains, A. Sharma for the design of the inactivating disulfide bond in Colla, and P. E. Milhier and his group for providing optimised PATRACK single-molecule tracking software. A.L.D. and M.C. thank T. Reddy for discussions and assistance with diffusion calculations. U.S. and C.G.B. thank the University of York Biology Electronic and Mechanical Workshops for custom fabrications, and G. Mashanov for custom image acquisition software. We also thank B. Berks for comments on the manuscript. C.G.B. acknowledges the Royal Society (2004/R1) and the University of York (Research Priming Fund) for financial support used to develop the TIRFM. P.R. acknowledges the late R. Saadia for his unwavering support. M.S.P.S. acknowledges access to the UK supercomputer ARCHER for molecular dynamics simulations. This work was supported by grants to C.K. and C.G.B. (BBSRC LoLa grant BB/G020671/1), J.P. (Deutsche Forschungsgemeinschaft SFB 944) and M.S.P.S. (BBSRC BB/L002558/1, Wellcome Trust WT092970MA).

Author Contributions P.R., O.B., J.P., M.S.P.S., C.G.B. and C.K. designed the experiments. N.A.C., assisted by C.G.B. and C.T., collected SMT-TIRFM data and C.T., assisted by N.A.C. and P.R., collected FRAP data for colicin-labelled OMPs using wild-type and deletion *E. coli* strains. P.R. conducted all ensemble TIRFM experiments on OMP islands. U.S. and C.G.B. designed and built the TIRF microscope at York used for all SMT-TIRFM experiments. O.B. and P.R. conducted all PSM experiments and data analysis. S.J.C. and C.G.B. designed and implemented software scripts used for diffusion simulations and bacterial cell image analysis. C.T., P.R., N.A.C., N.G.H. and R.K. purified all the proteins used in the study. C.T., P.R. and N.A.C. labelled colicins with fluorophores. D.M.Q. and T.J.G. constructed plasmids for colicin constructs and established procedures for growth of wild-type and gene deletion *E. coli* strains. M.C. and A.L.D. conducted the molecular dynamics simulations and subsequent analysis. C.K. drafted the paper with assistance from P.R., C.G.B., M.S.P.S. and J.P.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.K. (colin.kleanthous@bioch.ox.ac.uk) or C.G.B. (christoph.baumann@york.ac.uk)

METHODS

Confirmation of single gene deletions in Keio collection strains. All Keio mutants³¹ of *E. coli* BW25113 used in this work (with deletions in *lpp*, *ompA*, *pal*, *rfaC*, *tola* and *tonB*) were first validated by PCR, confirming the deletion strains contained the kanamycin resistance cassette within the open-reading frame of interest.

Colicin purification and fluorophore-labelling. Site-directed mutagenesis was used to introduce a solvent accessible cysteine (K469C) in the cytotoxic domain of a disulfide inactivated (Y324C, L447C) ColE9–Im9^{His6} construct, cloned into pET-21a (ref. 32). An equivalent Y324C, L447C, K469C ColE9 construct in which the N-terminal 83 amino acids of the colicin had been deleted (Δ^{1-83} ColE9) was also generated. This intrinsically disordered region of the colicin contains important protein–protein interaction epitopes that are required for translocation into the cell¹⁰. Removal of this region had no effect on the localization of OMP islands or on the turnover of OMPs. It was used primarily during long time courses where reduction of the inactivating disulfide bond was possible.

Standard recombinant molecular biology techniques were used to clone the *colla* gene into pET-21a with a carboxy-terminal His₆-tag and to introduce an inactivating disulfide within the coiled-coil R-domain (L257C, A411C), analogous to that used to inactivate ColE9, in addition to a solvent exposed cysteine in the cytotoxic domain (K544C).

All recombinant proteins were expressed in *E. coli* BL21 (DE3) cells and purified by nickel affinity and size-exclusion chromatography using published procedures³³. Elution from the nickel affinity column was achieved through guanidine denaturation of the ColE9–Im9^{His6} complex and imidazole elution of Colla^{His6}.

Cys469 and Cys544 within ColE9 and Colla, respectively, were labelled with a 20-fold molar excess of fluorophore (Alexa Fluor 488-maleimide (Invitrogen) or TMR-maleimide (Sigma)) in 20 mM potassium phosphate, pH 7.0, 2 M guanidine-HCl for 30 min at 37 °C followed by dialysis against 20 mM potassium phosphate, pH 7.0, 0.5 M NaCl at 4 °C. Dialysed proteins were further purified by size-exclusion chromatography (Superdex S200 HR 10/30, GE Healthcare) performed in the same buffer at room temperature and fractions containing fluorescently labelled protein analysed by SDS–PAGE, pooled, snap frozen and stored in aliquots at –20 °C. All labelling and subsequent purification steps were performed in the dark. The labelling efficiency (typically >0.7 fluorophores/protein) was estimated from the spectrophotometrically determined fluorophore (Alexa Fluor 488, $\epsilon_{\text{max}} = 71,000 \text{ cm}^{-1} \text{ M}^{-1}$; TMR, $\epsilon_{\text{max}} = 80,000 \text{ cm}^{-1} \text{ M}^{-1}$) and colicin (ColE9, $\epsilon_{280 \text{ nm}} = 46,075 \text{ cm}^{-1} \text{ M}^{-1}$; Colla, $\epsilon_{280 \text{ nm}} = 59,360 \text{ cm}^{-1} \text{ M}^{-1}$) concentrations after correcting for absorption at 280 nm by the fluorophore (Alexa Fluor 488, $A_{280 \text{ nm}} = 0.11 \times A_{495 \text{ nm}}$; TMR, $A_{280 \text{ nm}} = 0.3 \times A_{547 \text{ nm}}$). Analysis of single-molecule photobleaching characteristics for fluorophore-labelled colicins adsorbed on a quartz slide surface and viewed by TIRFM were consistent with labelling at a single position.

Membrane protein purification. OmpF and BtuB were purified according to previously published protocols^{10,12}. The purified proteins contained some endogenous LPS carried through the purification. An artificial transmembrane domain (ALAALAALAALAALAKSSR) was fused to the C terminus of maltose-binding protein (TM-MBP) by insertion of the corresponding oligonucleotide linker into the MCS of pMALc2x via restriction with BamHI and HindIII (ref. 34). For efficient purification and fluorescent labelling, an N-terminal His₁₀-tag and a ybbR-tag (DSLEFIASKLA)³⁵ were inserted via the NdeI site. This protein was expressed in *E. coli* TG-1 cells at 37 °C according to standard protocols. After cell lysis, the membrane fraction was solubilized with buffer containing 20 mM Triton X-100 and the protein purified by immobilized metal ion affinity chromatography (5 ml HiTrap Chelating, GE Healthcare). The protein was then covalently modified using a twofold molar excess of coenzyme A–Dy647 in the presence of 5 μM of the phosphopantetheinyl transferase Sfp and 10 mM Mg²⁺ for 1 h at room temperature³⁵. Finally, labelled TM-MBP (TM-MBP^{Dy647}) was further purified by size-exclusion chromatography in 20 mM Hepes, 150 mM NaCl, pH 7.5 containing 0.6 mM Triton X-100 (Superdex 200 10/300 GL, GE Healthcare).

Cell growth and OMP staining with colicins and antibodies. *E. coli* JM83 cells were grown at 37 °C in LB broth to exponential phase, after which 200 μl of cells were transferred to 4 ml M9-glucose minimal media (0.1 mM CaCl₂, 0.1 mM FeSO₄, 2 mM MgSO₄, 1 g l^{–1} NH₄Cl, 0.05% (w/v) casamino acids, 0.0002% (w/v) thiamine, 0.4% (w/v) D-glucose) and grown until absorbance (*A*) at 600 nm of ~0.4. Then 200 μl of cells were pelleted by centrifugation (4,700g, 3 min), with the pellet resuspended in 200 μl of fresh supplemented M9-glucose containing 300 nM fluorescently labelled colicin. After 15 min incubation at room temperature with mixing by rotary inversion, cells were washed twice by pelleting (4,700g, 3 min) and resuspended in 500 μl fresh M9-glucose, before finally resuspending the pellet in 100 μl supplemented M9-glucose. For sequential labelling of old and new OMPs, cells were grown for a further hour after the initial labelling such that

the *A*_{600 nm} had approximately doubled before a second staining was performed. Wash steps were repeated as above. For the transient expression of BtuB¹⁰ and TatA–GFP³⁶ from pBAD plasmids, cultures were grown in the presence of 100 $\mu\text{g ml}^{-1}$ ampicillin, and 0.2 mM L-arabinose added for 1 h before cells being used in confocal experiments.

For detection of BamA, JWD3 cells transformed with a pET17b plasmid expressing BamA–HA–L7 were grown in the presence of 100 $\mu\text{g ml}^{-1}$ ampicillin and 50 $\mu\text{g ml}^{-1}$ kanamycin, and prepared for microscopy as described for JM83 cells above. Insertion of an HA epitope into loop 7 of BamA was previously shown to be tolerated functionally and exposed on the external surface of bacteria¹⁹. BamA–HA–L7 was labelled *in vivo*, adding a 1:500 dilution of Alexa⁴⁸⁸-anti-HA antibody (Molecular Probes) and 1% (w/v) ultrapure BSA (Invitrogen) to 100 μl of JWD3 cells. After 15 min incubation with mixing by rotary inversion at room temperature, excess label was removed as described above for colicins.

For BamC labelling, a 1:1,000 dilution of mouse anti-BamC antibody²⁰ (gift from S. Buchanan) was added to 100 μl of JM83 cells prepared as above in the presence of 2% (w/v) ultrapure BSA and incubated for 15 min with mixing by rotary inversion at room temperature. Cells were then washed twice by pelleting (4,700g, 3 min) and resuspended in 500 μl of fresh supplemented M9 media, before a 1:500 dilution of Alexa⁴⁸⁸-anti-mouse secondary antibody (Molecular Probes) was added in the presence of 2% (w/v) ultrapure BSA. After 15 min incubation with rotary inversion at room temperature, excess label was removed as described above for colicins. For sequential labelling of old OMPs and new BamC, old BamC was first blocked with mouse anti-BamC antibody and unlabelled anti-mouse secondary antibody, as described above, while old BtuB or old Cir were stained with ColE9^{TMR} or Colla^{TMR}, respectively. After 1 h of growth at 37 °C in M9-glucose media, in which the *D*_{600 nm} had approximately doubled, new BamC was then labelled with mouse anti-BamC antibody and Alexa⁴⁸⁸-anti-mouse secondary antibody, implementing the same wash steps described above.

FRAP microscopy. For FRAP microscopy, 25 μl of stained *E. coli* cells were immobilised on poly-D-lysine (30–70 kDa, Sigma) coated coverslips and then imaged for less than 1 h at 20 or 37 °C. Cell viability under identical experimental conditions was verified using the LIVE/DEAD BacLight bacterial viability kit (Molecular Probes). Measurements were made using a Zeiss laser-scanning confocal microscope (LSM 510 Meta or LSM 710 Meta/Axiocvert 200M) equipped with 30 mW Ar ion and 1 mW HeNe lasers. Optical magnification was provided by a 63 \times oil-immersion objective (Zeiss, numerical aperture (NA) 1.4). Bleaching of fluorescently labelled proteins was performed using 20 (488 nm laser) or 500 (543 nm laser) scan iterations (1.58 μs pixel dwell time) over a rectangular region of interest with a 10 \times digital zoom and the laser power set to 50–100% (depending on laser line). To minimise photobleaching, FRAP was monitored using reduced laser power (2–10% depending on laser line). The diameter of the pinhole was varied between 0.2 and 0.3 μm depending on the sample conditions and the fluorophore used. Two-colour FRAP experiments (GFP and TMR) on a single bacterial cell were performed as follows: (1) acquire pre-bleach and first post-bleach image with 543 nm laser illumination, (2) record complete FRAP sequence for TatA–GFP with 488 nm laser illumination, and (3) record post-bleach images with 543 nm laser illumination. The individual laser sources were used separately to prevent cross-talk between the two fluorescence emission channels. A DIC microscopy image was acquired before and after each FRAP experiment to ensure the image of the bacterial cell remained in focus and adherent to the PDL coated surface. In all cases, DIC images were recorded using a 633 nm laser to prevent photobleaching of fluorescent probes.

In vivo SMT–TIRFM. JM83 cells were immobilised on poly-D-lysine coated quartz slides (25 mm \times 75 mm, cleaned in 1 M KOH) in supplemented M9 media containing 100 nM fluorescently labelled colicin. An ultra-thin sample chamber was formed by adding ~0.1 $\mu\text{g ml}^{-1}$ of 5 μm silica beads to the media, overlaying with a no. 1 coverslip (22 mm \times 64 mm) and sealing with nail varnish. An inverted Zeiss IM35 microscope chassis fitted with a 100 \times oil-immersion objective (Zeiss Plan-Apochromat, NA 1.4) was modified in-house for prism-coupled TIRFM. Illumination was provided by 488 nm (30 mW) and 561 nm (30 mW) optically pumped solid-state lasers (Sapphire LP, Coherent). A $\lambda/4$ waveplate was used to depolarize the laser excitation, thus ensuring fluorophore orientation effects were minimised. The fluorescence image was separated into two channels using an image splitter (Optosplit II, Cairn Research) with appropriate dichroic beamsplitter (FT580, Zeiss) and band pass filters for Alexa Fluor 488 (ET525/50M, Chroma) and TMR (HQ605/20M, Chroma). Background data sets were obtained by directly adsorbing fluorescently-labelled colicins to poly-D-lysine-coated quartz slides. Moderate photobleaching of each sample was done to enable detection and tracking of single fluorophores. Analogue video (IC-300B intensified CCD, Photon Technology International) was recorded at 30 frames per second (f.p.s.) (494 \times 786 pixels²) and digital video (DU897E emCCD, Andor; Evolve 512 emCCD, Photometrics) was recorded at 30 f.p.s. (512 \times 512 pixels²) or 58 f.p.s.

(128 × 128 pixels). All video data was collected at room temperature (20–22 °C). In analogue video data, single fluorophores were identified manually and tracked using GMimPro software³⁷. In digital video data, single fluorophores were identified automatically (using standard full-width at half maximum (FWHM) and threshold values) and tracked using the automated SPT function in GMimPro. *D* values were extracted from the time-dependence of the MSD using either linear regression or power law fitting of the initial 4–5 time delays. The confinement diameter (*d*) was calculated according to the equation: $AMV = (d/2)^2/6$, in which AMV is the asymptotic MSD value³⁸.

Monte Carlo simulations of two-dimensional diffusion. MATLAB was used to model diffusion in the *E. coli* outer membrane using a standard Monte Carlo approach³⁸. At each time point (*dt*), a diffusion step (*dl*) with length $(4Ddt)^{-1/2}$, in which *D* is the diffusion coefficient, was taken in a uniformly distributed random direction^{39,40}. All simulations were performed for 5,000 particles, over a total simulation time (*t*) of 2 s at intervals $dt = 1 \mu s$, unless otherwise stated. Particle runs were performed sequentially; therefore, to reduce computational overheads, the cumulative moving mean (\bar{x}_n) and variance (σ_n^2) of the MSD were calculated following each particle run, where *n* is the current iteration:

$$\bar{x}_n = \frac{(\sum_{i=1}^{n-1} x_i + x_n)}{n}$$

$$\sigma_n^2 = \frac{(n-1)\sigma_{n-1}^2 + (\bar{x}_n - \bar{x}_{n-1})(x_n - \bar{x}_n)}{n}$$

Curvature of the cellular membrane was modelled by mapping planar diffusion coordinates onto a hemispherically capped cylinder of length 3 μm and radius 250 nm. The limited TIRF evanescent field depth was modelled by terminating, but not discarding, particle paths which diffused further than 150 nm from the closest observable plane.

Molecular crowding in the outer membrane was simulated using excluded volume of BtuB and two of the most abundant proteins in the outer membrane, OmpA and OmpF. Each protein species was modelled in 2D as a series of single, non-overlapping, hard circles with radii of $R_{OmpA} = 1.3$ nm, $R_{OmpF} = 3.1$ nm and $R_{BtuB} = 2.3$ nm, and with relative populations of 333, 111 and 1, respectively. A random planar distribution of OmpA and OmpF was generated for each diffusing BtuB up to the percentage membrane occupancy, δ . BtuBs were randomly positioned and subject to the standard Monte-Carlo-based diffusion approach; however, only steps ending in vacant positions (no overlap of BtuB with OmpA or OmpF) were accepted. To model association of BtuB with the OmpF population in the outer membrane, the BtuB radius was increased randomly in a single step according to the following equation and association rate, *k*.

$$R_{BtuB+OmpF} = \sqrt{R_{BtuB}^2 + R_{OmpF}^2}$$

Planar diffusion trajectories were subsequently mapped to the curved bacterial surface as described previously. For optimisation of the MSD simulated under molecular crowding conditions, the rate *k* and parameters *D* ($= 0.01$ – $0.3 \mu m^2 s^{-1}$) and δ ($\leq 20\%$) were varied manually. Very similar parameters to those for OmpF–BtuB association were generated for OmpA–BtuB association, thus only those for OmpF are shown in Extended Data Fig. 2c. Monte Carlo simulations that did not include promiscuous association between OMPs, but merely membrane crowding ($\delta \leq 20\%$), did not reproduce the *in vivo* diffusive behaviour (data not shown), which is consistent with an increase in mass of the tracked diffusing species explaining the asymptotic experimental MSD observed *in vivo* and *in vitro*.

Analysis of OMP islands by TIRFM. JM83 cells were immobilised on poly-D-lysine and viewed using prism-coupled TIRFM as described above. 100 consecutive video frames were stacked to build an image that was then analysed. Only sequences with no saturation of the intensified CCD camera were selected. For each picture and each channel, we used ImageJ⁴¹ software to select only pixels that were three times more intense than background for further quantification. Intensity quantification within individual spots was obtained using ImageJ. Within each cluster or island, the total pixel intensity was normalized so the population corresponding to a single photobleaching step was fitted to 1 arbitrary unit (AU).

CoLocalizer Pro 2.7.1 software (CoLocalization Research Software, <http://www.colocalizer.com>)⁴² was used for co-localization of red and green channels. The co-localization value corresponded to the ratio between yellow pixels over the sum of yellow, red and green pixels on each overlapped picture. Distribution of fluorescence intensity along the *x* axis of bacteria was determined using a custom script implemented in MATLAB (version 2012a, MathWorks). Raw images were initially thresholded against a user-defined intensity (selection via a graphical user interface) to provide a binary image, approximately highlighting bacteria against

the background. High frequency noise was reduced through application of the built-in MATLAB medfilt2 median filter. Individual bacteria were identified as continuous high intensity pixel regions, with pixels of a continuous region required to be within a $0.6 \mu m^2$ area of each other. Further filtering was performed to remove bacteria less than 1 μm away from their neighbour to improve the quality of intensity profile analysis results. Bacterial poles were identified as corresponding to the two pixels in a continuous region with the largest separation. Intensity along the selected bacteria was measured 11 times between the bacterial end-points, each at a uniformly-spaced offset from the bacterial long axis in the range ± 200 nm. The mean profile is calculated and normalized to the range 0–100 for comparison between bacteria.

Confocal time-lapse microscopy. For confocal microscopy, 200 μl of M9 containing 1% agarose (w/v) was introduced into a Gene Frame matrix that was previously adhered to a clean slide. The agar pad was formed by addition of a clean coverslip on top until solidification had occurred. Then, 10 μl of stained bacteria was added to the pad, which was sealed afterward using a clean coverslip. Cell growth was monitored at 37 °C and visualized through 100× oil-immersion objective (Zeiss Plan-APOCHROMAT NA 1.4) on a Zeiss LSM780 microscope. Individual cells were selected in a region of interest that was clear of other cells. Snapshots were taken every 30 min during 2–3 h time courses, using scan speed 7 and 1–2% of laser power (depending on laser line).

Formation of polymer-supported membranes. Modification of glass coverslips was performed as described previously³⁴. In brief, surfaces were first cleaned by plasma cleaning and then activated by silanization with pure (3-glycidyloxypropyl)trimethoxysilane (Sigma) at 75 °C for 1 h. Subsequently, the surfaces were reacted with molten bis-amino-polyethylene glycol (PEG) with a molecular mass of 2,000 daltons (Da) (Rapp Polymere) for 4 h at 75 °C. Modification of the free amines with palmitic acid (Sigma) was carried out in the presence of an excess of diisopropylcarbodiimide (Sigma). Formation of proteoliposomes was carried out by detergent extraction in lipid-protein-detergent mixtures by addition of β-cyclodextrin⁴³. Lipids, proteins and lipopolysaccharides solubilized in Triton-X-100 or *n*-octyl-β-D-glucoside were combined as required before a twofold excess of heptakis(2,6-di-*O*-methyl)-β-cyclodextrin (Sigma) over the detergent was added. Reconstitution of BtuB was achieved by mixing an *E. coli* lipid extract with the protein in MES buffer pH 6.5 and adding a twofold molar excess of β-cyclodextrin. Experiments were conducted using either a defined *E. coli* polar lipid extract (PE 67%; PG 23.2%; cardiolipin (CA) 9.8%) or total *E. coli* lipid extract (PE 57.5%; PG 15.1%; CA 9.8%; unknown 17.6%) from Avanti Lipids. OMP mobilities in PSM experiments were essentially identical using these lipid extracts. So as to obtain a 1× final concentration of protein, a molar ratio of 1:100,000 (protein:lipid) was prepared and vortexed for 2 s before deposition on surfaces. After 10 min the proteoliposomes were added onto functionalized surfaces and incubated for 20 min for immobilization. The buffer was then exchanged for a solution of 10% (w/v) PEG with a molecular mass of 8,000 Da (Sigma Aldrich) to induce fusion of the proteoliposomes into a homogenous lipid bilayer. After 20 min, the surfaces were extensively washed with buffer to remove excess vesicles. Before imaging, bilayer continuity was visualized by monitoring the homogeneous lateral diffusion of the lipids DiD^{Dy647} or LPS^{BODIPY} in the membrane. Before data acquisition, BtuB was labelled by the addition of 100 nM ColE9^{TMR} for 5 min, followed by removal of unbound ligand. Assuming all BtuB molecules were labelled, and single step photobleaching corresponded to single molecules, the density of BtuB at 1× concentration was comparable to that estimated *in vivo* (~ 50 molecules per μm^2).

Tracking single molecules in polymer supported bilayers. OMPs were tracked using TIRF microscopy carried out with an Olympus IX71 inverted microscope equipped with a quad-line total internal reflection illumination (TIR) condenser (Olympus), a 150× magnification objective with a numerical aperture of 1.45 (UAP0 150×/1.45 TIRFM, Olympus) and a back-illuminated emCCD camera (Andor iXon Ultra 897). TMR was illuminated by a 561-nm diode-pumped solid-state laser (Cobold Jive, 200 mW) with 2 mW output power at the objective, whereas BODIPY and Dy647 were illuminated with 488 and 642 nm laser diodes (Omicron, 140 and 200 mW, respectively) with output powers of 5 and 4 mW, respectively. Fluorescence was detected using a quadband emission filter (446/523/600/677 HC Quadband Filter, Semrock) and recorded at 30 f.p.s. All experiments were carried out using buffer complemented with oxygen scavenger and a redox-active photoprotectant (0.5 mg ml⁻¹ glucose oxidase (Sigma), 0.04 mg ml⁻¹ catalase (Roche Applied Science), 5% glucose (w/v), 1 μM ascorbic acid and 1 μM methyl viologene) to minimize photobleaching⁴⁴. Moderate bleaching of the sample was applied to be able to track single molecules.

All videos were analysed using custom software (PATRACK), implemented in visual C++ and provided by P.-E. Milhiet and P. Dosset. The centre of each fluorescence peak was determined with sub-pixel resolution by fitting a

two-dimensional elliptical Gaussian function. The two-dimensional trajectories of single molecules were constructed frame-by-frame, selecting particles that displayed a single bleaching step. Diffusion coefficient values were determined from a linear fit to the MSD plots between the first and fourth points according to the equation: $\text{MSD}(t) = 4Dt$.

We used a new algorithm within PATRACK based on a back-propagation neural network allowing automatic detection of Brownian, confined and directed motion modes within a trajectory. The ability of the software to accurately detect different diffusion modes was first established on simulated data and then evaluated on trajectories recorded from fluorescently labelled molecules in two different biological systems^{45,46}.

Statistical analysis of microscopy data. All experiments were conducted at least twice (no significant differences were observed between data sets either by Student's *t*-test or non-parametric Mann–Whitney test) and in each case a representative set of data presented in the figures. Sampling sizes in experiments (cells, OMP islands, single molecule trajectories) were validated by very significant *P* values in a student *t*-test or a non-parametric Mann–Whitney test. In the analysis of OMP islands in post growth conditions by TIRFM, cells that were undergoing division were explicitly selected for imaging. No statistical methods were used to predetermine sample size.

Molecular dynamics simulations. Simulations were run using GROMACS (version 4) (<http://www.gromacs.org>) using a modified version of the MARTINI forcefield^{47–49}. Structures of OmpF (PDB code 2OMF) and BtuB (PDB code 2GUF) were converted to coarse-grained models as described previously⁴⁹. In the model, each coarse-grained particle represents approximately four atoms in the atomistic structure. Protein molecules were represented with a single coarse-grained particle per backbone moiety, with variable numbers of sidechain particles. To maintain the β -barrel structure of OmpF and BtuB, an elastic network model was used, with a force constant $1,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ applied between C α particles within 7 \AA of one another.

To mimic varied levels of crowding in the membrane, several systems of varying size and protein content were constructed (see Extended Data Fig. 6). Proteins were inserted into a pre-equilibrated bilayer containing POPE and POPG in a 3:1 ratio. The lipid composition was selected to approximate that used in PSM experiments. Using a membrane patch of $\sim 30 \times 30 \times 10 \text{ nm}^3$, one of the following protein groups was inserted: a single OmpF trimer; a single BtuB molecule; two OmpF trimers and two BtuB monomers or 4 BtuB monomers; four OmpF and five BtuB or 9 BtuB monomers; or eight OmpF and eight BtuB. Into a larger membrane patch of $\sim 60 \times 60 \times 10 \text{ nm}^3$ were inserted: eight OmpF and eight BtuB; or 18 OmpF and 18 BtuB or 36 BtuBs (see Fig. 4c and Extended Data Fig. 6 to see final snapshots of the systems). Large systems were simulated to identify how system size might influence results, since differences have previously been observed⁵⁰. In each case, OmpF and BtuB molecules were positioned on a square grid in an alternating pattern. To ensure no protein interaction face was favoured in the starting configuration we randomized the rotation of each protein around its centre of mass about the membrane normal. Systems were solvated and sodium counter ions were added to neutralize the system. Each system was equilibrated for 100 ns with the proteins restrained in the *X–Y* plane, using a force constant of $1,000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$. Simulations were performed using a time step of 20 fs. A temperature of 313 K was maintained using the Berendsen thermostat, and coupled separately for proteins, lipids, and solvent. Pressure was maintained at 1 bar with a semi-isotropic coupling and using the Berendsen barostat, with a compressibility of $5 \times 10^{-6} \text{ bar}^{-1}$. Electrostatic interactions were shifted between 0.0 and 1.2 nm, and Lennard–Jones interactions were shifted between 0.9 and 1.2 nm. Production runs, with no restraints applied, were for 10 μs (see Extended Data Fig. 6).

Molecular dynamics simulations analysis. Analysis was performed using in house scripts, MDAnalysis⁵¹, and VMD⁵².

Protein diffusion coefficients were calculated using an in-house python script (<http://dx.doi.org/10.5281/zenodo.11827>). In brief, MSD was calculated for time windows between 1 ns and 10 % of the full simulation time (that is, 1 μs), to allow for suitably sampled windows. Parameters were obtained by a linear fit to the MSD versus time data, with error estimated as the difference in the slopes of the two halves of the data.

To quantify interaction frequencies occurring during the simulations, we first define an interaction count function, $C_{\sigma,t}(\text{res}_{i,p_1}, \text{res}_{j,p_2})$, to perform a simple count of the interactions between residues *i* and *j* of protein *p*₁ and *p*₂, respectively, at time *t* in simulation σ :

$$C_{\sigma,t}(\text{res}_{i,p_1}, \text{res}_{j,p_2}) = \begin{cases} 1 & \text{if the centroids of res}_i \text{ and res}_j \text{ are within } 8 \text{ \AA} \text{ of each other} \\ 0 & \text{otherwise} \end{cases}$$

Thus, the interaction frequency, $\text{int}(\text{res}_i, \text{res}_j)$, between any two residues, *res*_{*i*} in BtuB and *res*_{*j*} in OmpF, is given by:

$$\text{int}(\text{res}_i, \text{res}_j) = \sum_{\sigma=1}^{N_{\text{sim}}} \sum_{t=T-100}^T \sum_{p_1=1}^{N_{\text{BtuB},\sigma}} \sum_{p_2=1}^{N_{\text{OmpF},\sigma}} C_{\sigma,t}(\text{res}_{i,p_1}, \text{res}_{j,p_2})$$

where N_{sim} is the total number of simulations being considered; *T* is the total simulation time (in ns); and $N_{\text{BtuB},\sigma}$ and $N_{\text{OmpF},\sigma}$ are the total number of BtuB and OmpF proteins, respectively, in simulation σ . Only the last 100 ns of each simulation was used. The proportional interaction frequency, $I_{i,j}$, is then given by normalizing the interaction frequency for any pair of residues, $\text{int}(\text{res}_i, \text{res}_j)$ by the total number of interactions:

$$I_{i,j} = \frac{\text{int}(\text{res}_i, \text{res}_j)}{\sum_{i=1}^{N_{\text{res BtuB}}} \sum_{j=1}^{N_{\text{res OmpF}}} \text{int}(\text{res}_i, \text{res}_j)}$$

in which $N_{\text{res BtuB}}$ and $N_{\text{res OmpF}}$ denote the number of residues in a BtuB and OmpF single protein, respectively.

To look at interactions of BtuB with itself, $\text{int}(\text{res}_i, \text{res}_j)$ is instead defined by:

$$\text{int}(\text{res}_i, \text{res}_j) = \sum_{\sigma=1}^{N_{\text{sim}}} \sum_{t=T-100}^T \sum_{p_1=1, p_1 \neq p_2}^{N_{\text{BtuB},\sigma}} \sum_{p_2=1}^{N_{\text{BtuB},\sigma}} \frac{1}{2} C_{\sigma,t}(\text{res}_{i,p_1}, \text{res}_{j,p_2})$$

The factor of one half accounts for each interaction being counted twice. Similar to BtuB–OmpF interactions, $\text{int}(\text{res}_i, \text{res}_j)$ is normalized by the total number of interactions to give the proportional interaction frequency, $I_{i,j}$ for any two residues, *res*_{*i*} and *res*_{*j*}, on different monomers of BtuB. The proportional interaction frequency, I_i of a single residue, *res*_{*i*}, of BtuB with OmpF—the values shown in red bars on the side of the interaction frequency matrix plots (Extended Data Fig. 7)—is then given by summing the proportional interaction frequencies of *res*_{*i*} over all residues of OmpF:

$$I_i = \sum_{j=1}^{N_{\text{res OmpF}}} I_{i,j} = \frac{\sum_{j=1}^{N_{\text{res OmpF}}} \text{int}(\text{res}_i, \text{res}_j)}{\sum_{i=1}^{N_{\text{res BtuB}}} \sum_{j=1}^{N_{\text{res OmpF}}} \text{int}(\text{res}_i, \text{res}_j)}$$

Similarly, a single residue function is defined for the proportional interaction frequencies of OmpF with BtuB, and residues of BtuB in their interaction with other BtuB monomers.

Amino acid interaction propensities: to assess the involvement of each amino acid in mediating BtuB–OmpF and BtuB–BtuB interactions, a residue interaction propensity is defined, adapted from the residue propensity metric defined previously⁵³ in the hallmark work on protein–protein interactions. For each amino acid, AA_{*k*}, the interaction propensity, IP_{AA_k} , for the interaction of residues of BtuB with OmpF is given by comparing the proportion of BtuB interactions mediated by amino acid AA_{*k*} with the proportion of the BtuB surface it represents:

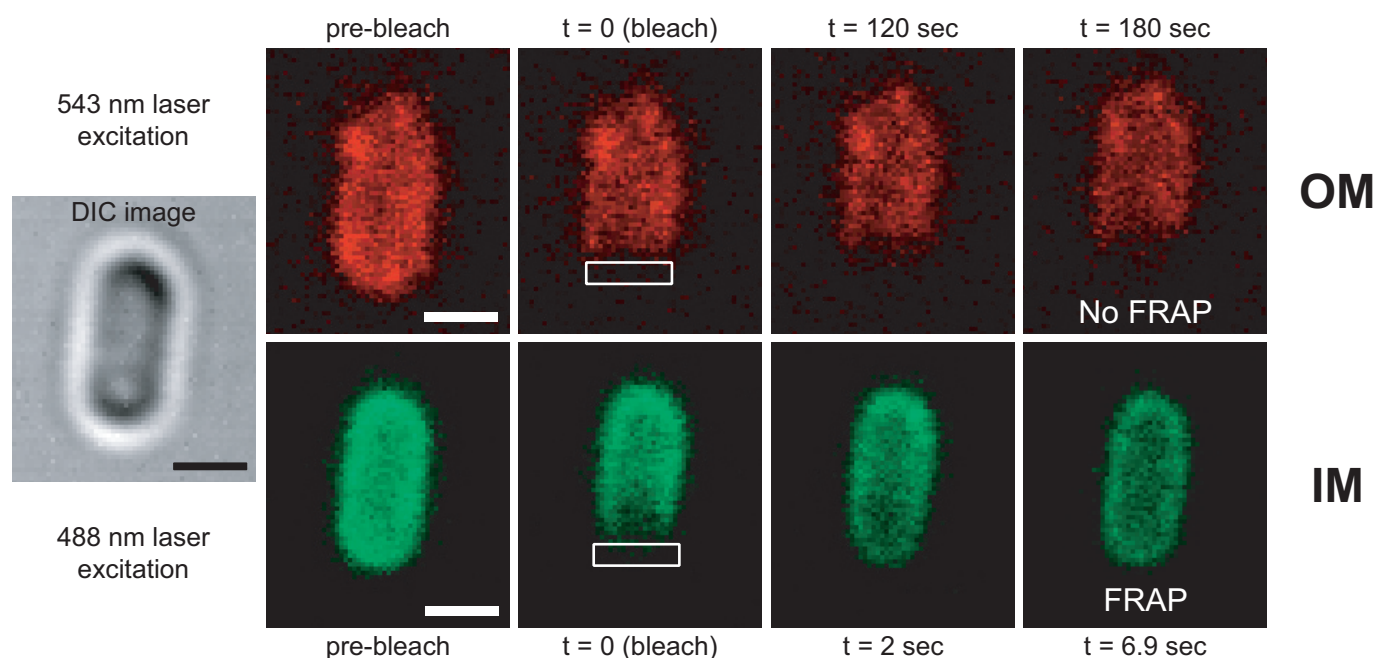
$$\text{IP}_{\text{AA}_k} = \frac{\left(\sum_{i=\text{AA}_k} I_i \right) / \left(\sum_{i=1}^{N_{\text{res BtuB}}} I_i \right)}{N_{\text{AA}_k, \text{BtuBsurf}} / N_{\text{res BtuBsurf}}}$$

where $N_{\text{AA}_k, \text{BtuBsurf}}$ is the number of amino acid AA_{*k*} on the surface of BtuB, and $N_{\text{res BtuBsurf}}$ is the total number of residues on the surface of BtuB. Similarly, residue interaction propensities are defined for the OmpF interaction with BtuB and interaction of BtuB with other monomers of BtuB.

- Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
- Penfold, C. N. *et al.* Flexibility in the receptor-binding domain of the enzymatic colicin E9 is required for toxicity against *Escherichia coli* cells. *J. Bacteriol.* **186**, 4520–4527 (2004).
- Garinot-Schneider, C., Pommer, A. J., Moore, G. R., Kleinhous, C. & James, R. Identification of putative active-site residues in the DNase domain of colicin E9 by random mutagenesis. *J. Mol. Biol.* **260**, 731–742 (1996).
- Roder, F., Birkholz, O., Beutel, O., Paterok, D. & Piehler, J. Spatial organization of lipid phases in micropatterned polymer-supported membranes. *J. Am. Chem. Soc.* **135**, 1189–1192 (2013).
- Yin, J. *et al.* Genetically encoded short peptide tag for versatile protein labeling by Sfp phosphopantetheinyl transferase. *Proc. Natl Acad. Sci. USA* **102**, 15815–15820 (2005).
- Ray, N., Nieninger, A., Mullineaux, C. W. & Robinson, C. Location and mobility of twin arginine translocase subunits in the *Escherichia coli* plasma membrane. *J. Biol. Chem.* **280**, 17961–17968 (2005).
- Mashanov, G. I. & Molloy, J. E. Automatic detection of single fluorophores in live cells. *Biophys. J.* **92**, 2199–2211 (2007).
- Kusumi, A., Sako, Y. & Yamamoto, M. Confined lateral diffusion of membrane receptors as studied by single particle tracking (nanovid microscopy). Effects of calcium-induced differentiation in cultured epithelial cells. *Biophys. J.* **65**, 2021–2040 (1993).

39. Deich, J., Judd, E. M., McAdams, H. H. & Moerner, W. E. Visualization of the movement of single histidine kinase molecules in live *Caulobacter* cells. *Proc. Natl Acad. Sci. USA* **101**, 15921–15926 (2004).
40. Leake, M. C. *et al.* Stoichiometry and turnover in single, functioning membrane protein complexes. *Nature* **443**, 355–358 (2006).
41. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).
42. Zinchuk, V., Wu, Y., Grossenbacher-Zinchuk, O. & Stefani, E. Quantifying spatial correlations of fluorescent markers using enhanced background reduction with protein proximity index and correlation coefficient estimations. *Nature Protocols* **6**, 1554–1567 (2011).
43. Degrip, W. J., Vanoostrum, J. & Bovee-Geurts, P. H. Selective detergent-extraction from mixed detergent/lipid/protein micelles, using cyclodextrin inclusion compounds: a novel generic approach for the preparation of proteoliposomes. *Biochem. J.* **330**, 667–674 (1998).
44. Vogelsang, J. *et al.* A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chem.* **47**, 5465–5469 (2008).
45. Espenel, C. *et al.* Single-molecule analysis of CD9 dynamics and partitioning reveals multiple modes of interaction in the tetraspanin web. *J. Cell Biol.* **182**, 765–776 (2008).
46. Krementsov, D. N. *et al.* HIV-1 assembly differentially alters dynamics and partitioning of tetraspanins and raft components. *Traffic* **11**, 1401–1414 (2010).
47. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
48. Bond, P. J., Wee, C. L. & Sansom, M. S. Coarse-grained molecular dynamics simulations of the energetics of helix insertion into a lipid bilayer. *Biochemistry* **47**, 11321–11331 (2008).
49. Bond, P. J. & Sansom, M. S. Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.* **128**, 2697–2704 (2006).
50. Klauda, J. B., Brooks, B. R. & Pastor, R. W. Dynamical motions of lipids and a finite size effect in simulations of bilayers. *J. Chem. Phys.* **125**, 144710 (2006).
51. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, (2011).
52. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
53. Jones, S. & Thornton, J. M. Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA* **93**, 13–20 (1996).
54. Javanainen, M. *et al.* Anomalous and normal diffusion of proteins and lipids in crowded lipid membranes. *Faraday Disc.* **161**, 397–417 (2013).
55. Oddershede, L., Dreyer, J. K., Grego, S., Brown, S. & Berg-Sorensen, K. The motion of a single molecule, the lambda-receptor, in the bacterial outer membrane. *Biophys. J.* **83**, 3152–3161 (2002).
56. Spector, J. *et al.* Mobility of BtuB and OmpF in the *Escherichia coli* outer membrane: implications for dynamic formation of a translocon complex. *Biophys. J.* **99**, 3880–3886 (2010).
57. Leake, M. C. *et al.* Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by *in vivo* single-molecule imaging. *Proc. Natl Acad. Sci. USA* **105**, 15376–15381 (2008).
58. Jeanteur, D. *et al.* Structural and functional alterations of a colicin-resistant mutant of OmpF porin from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **91**, 10675–10679 (1994).

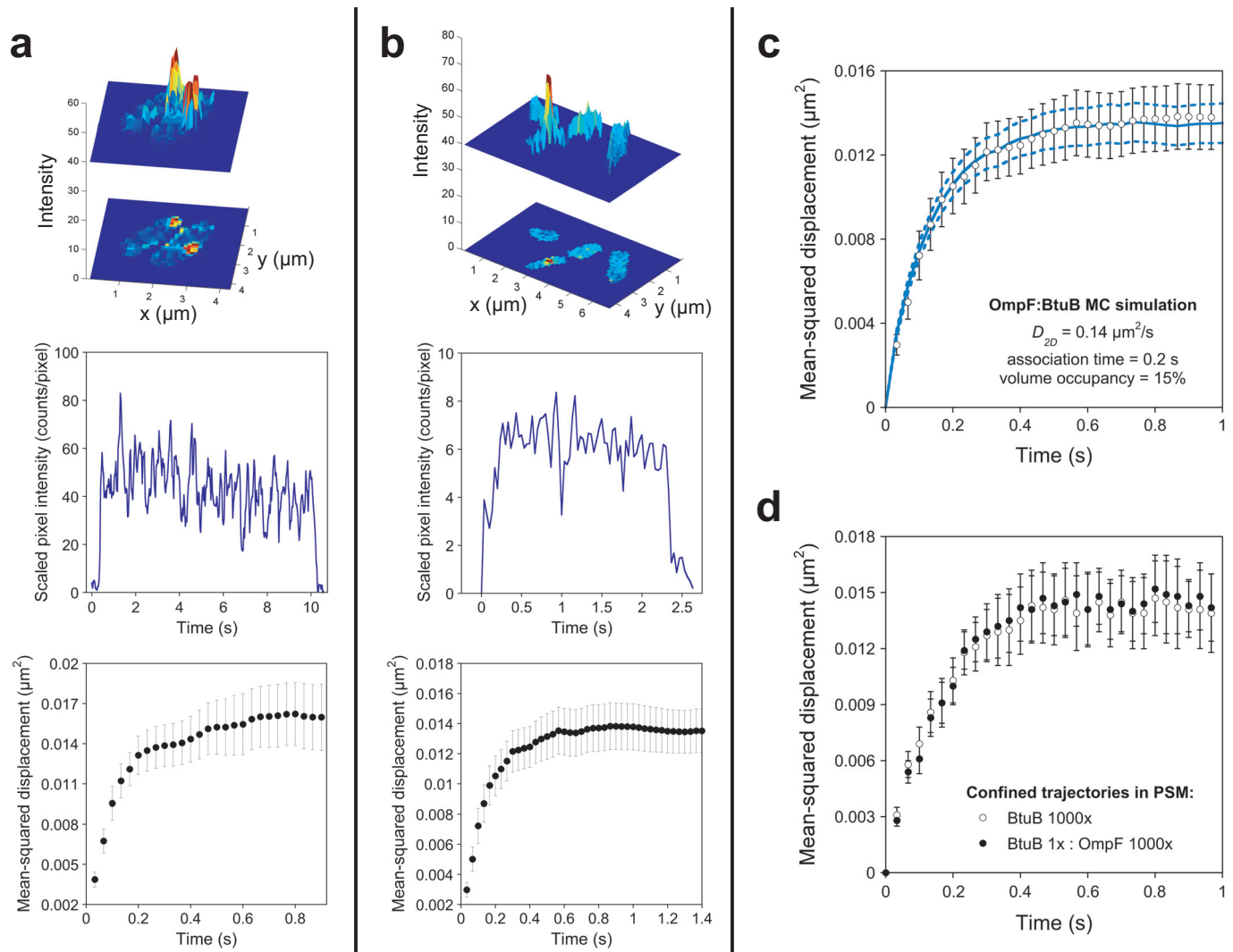
BtuB-ColE9^{TMR}



TatA-GFP

Extended Data Figure 1 | Visualizing the different diffusive behaviours of proteins in the inner and outer membranes of the same *E. coli* cell. Single *E. coli* JM83 bacterial cell visualized by laser-scanning confocal fluorescence microscopy at 37 °C. At the outer membrane (top row) BtuB labelled with ColE9^{TMR} showed no fluorescence recovery after photobleaching a region of interest with a 543 nm laser (rectangle). By contrast, recovery of TatA-GFP

fluorescence in the inner membrane of the same cell (bottom row) was observed within a few seconds after photobleaching an identical region with 488 nm laser. The unrestricted mobility of TatA-GFP in the inner membrane accounts for its analogue distribution during cell division (main text, Fig. 1b, f). This experiment was done in duplicate (one representative set of images is presented in the figure). Scale bar, 1 μ m.



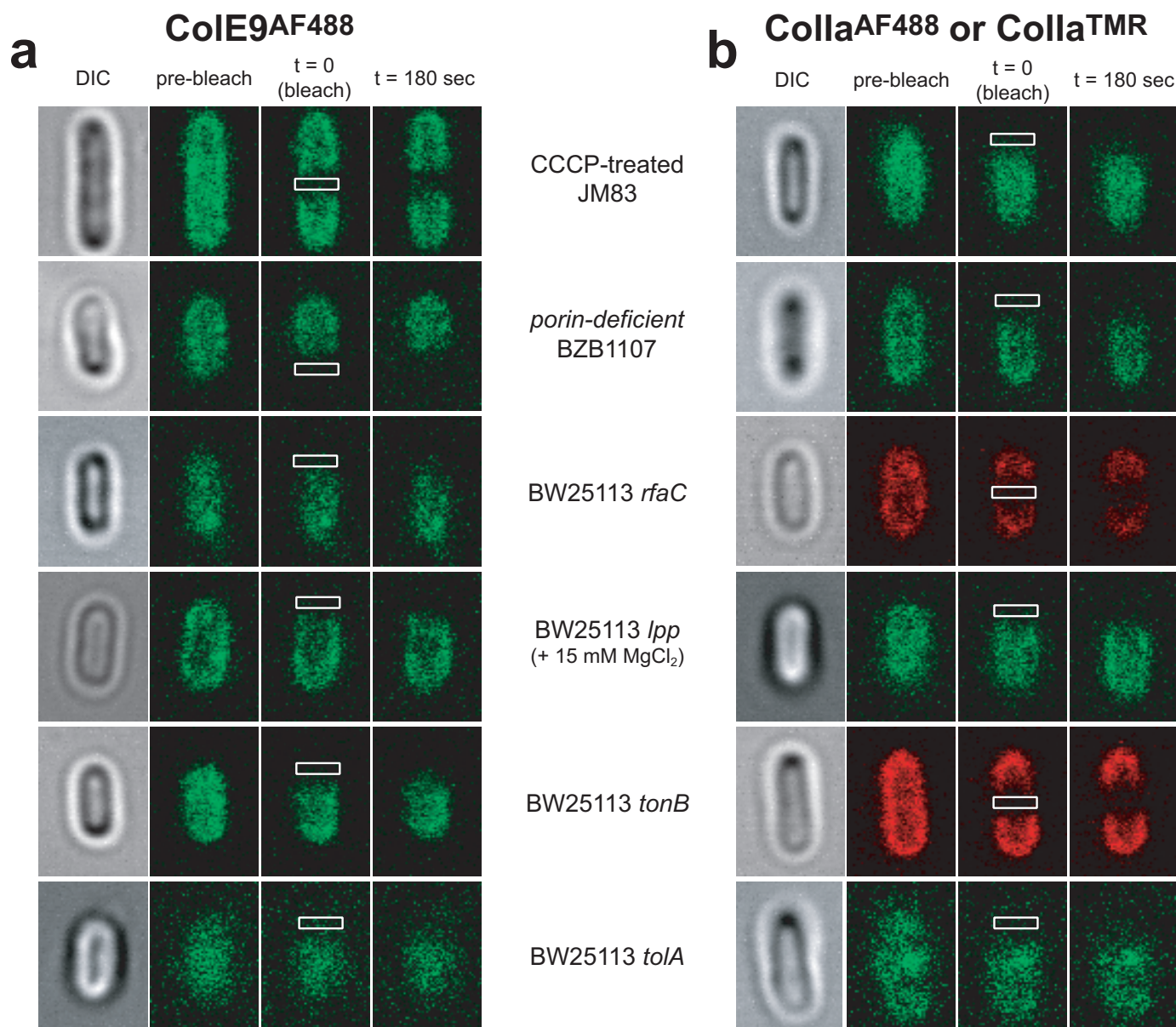
Extended Data Figure 2 | SMT-TIRFM defines the mobility of colicin-labelled BtuB and Cir *in vivo*. **a**, Top, Z-projection of fluorescence intensity (increasing from blue to red) of two Cir-CollA^{AF488} complexes on separate JM83 cells. Middle, the fluorescence intensity in consecutive images (30 f.p.s.) is displayed for a typical Cir-CollA^{AF488} fluorescent spot. The intensity and single-step photobleaching behaviour were consistent with a single membrane complex being tracked. Bottom, MSD was calculated for single Cir-CollA^{AF488} complexes ($n = 41$) that displayed single-step photobleaching behaviour, could be tracked for at least 1.7 s before photobleaching (error is reported as s.e.m.) and were not immobilized on the quartz surface ($\text{MSD}_{\text{end of trajectory}} > 0.004$, determined for surface bound CollA^{AF488} in the same samples). The MSD value rapidly approached an asymptotic value that was consistent with restricted lateral diffusion. Linear regression of the MSD for the first 4 time delays (0.033–0.13 s) yielded the planar diffusion coefficient ($D \approx 0.019 \mu\text{m}^2 \text{s}^{-1}$).

b, Top, Z-projection of fluorescence intensity (increasing from blue to red) of a BtuB-ColE9^{AF488} complex on a JM83 cell. Middle, the fluorescence intensity in consecutive images (30 f.p.s.) is displayed for a typical fluorescent spot. The intensity and single-step photobleaching behaviour were consistent with a single membrane complex being tracked. Bottom, MSD was calculated for individual BtuB-ColE9^{AF488} complexes ($n = 62$) that displayed single-step photobleaching behaviour, could be tracked for at least 1.7 s before photobleaching (error is reported as s.e.m.) and were not immobilized on the quartz surface ($\text{MSD}_{\text{end of trajectory}} > 0.008$, determined for surface bound

ColE9^{AF488} in the same samples). The MSD value rapidly approached an asymptotic value that was consistent with restricted lateral diffusion. Linear regression of the MSD for the first five time delays (0.033–0.17 s) yielded the planar diffusion coefficient ($D \approx 0.013 \mu\text{m}^2 \text{s}^{-1}$). See Extended Data Table 1 for all fitted values (from a minimum of four experimental replicates).

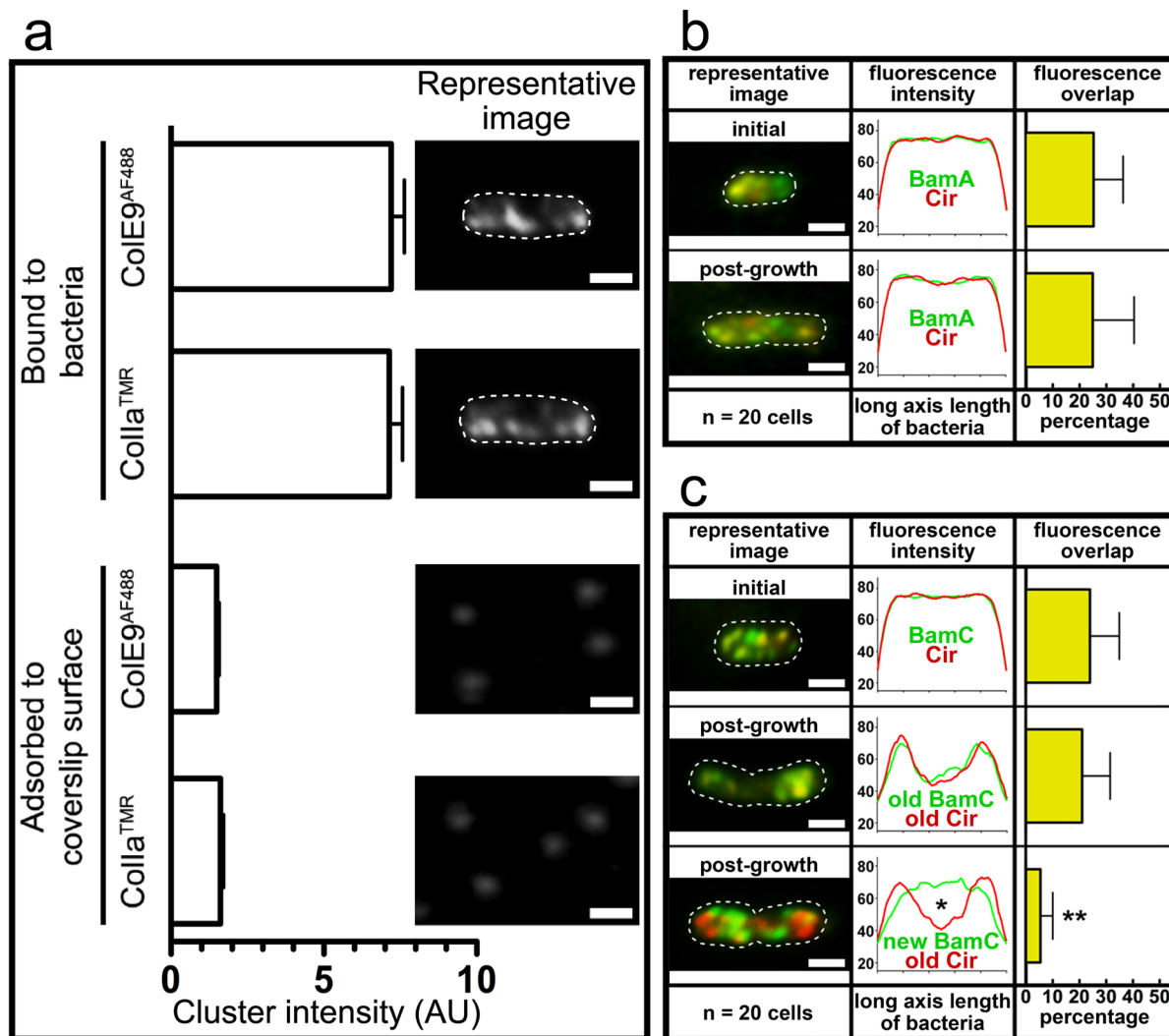
c, Comparison of experimental MSD for BtuB-ColE9^{AF488} complexes from **b** (open circles) with mean output from Monte Carlo simulations (solid blue line) of 2D diffusion ($D = 0.14 \mu\text{m}^2 \text{s}^{-1}$, association time of 0.2 s (k^{-1}), and 15% volume occupancy) in a curved, crowded membrane typical of a rod-shaped bacterium, which is illuminated by an evanescent field (penetration depth = 150 nm). Here lateral D was identical to the value observed for monomeric BtuB diffusing in PSMs (Extended Data Fig. 5c). The lower apparent lateral D measured experimentally ($= 0.013 \mu\text{m}^2 \text{s}^{-1}$) indicates the promiscuous PPIs induce temporal corralling. The dashed (blue) lines are the upper and lower 99.75% confidence limits for the mean ($n = 5,000$ trajectories). See Methods for details of the Monte Carlo simulations.

d, MSD plots (\pm s.e.m.) for all confined trajectories (500 in each case) of BtuB-ColE9^{TMR} complexes diffusing in PSMs, red data points in Fig. 4b (panels 3 and 4), main text. [BtuB] \times 1,000 data (open circles) is compared to data for native levels of BtuB (closed circles) in the presence of native levels of OmpF (equivalent to [BtuB] \times 1000). Fits to both sets of data yield values for D and confinement diameter, $\sim 0.02 \mu\text{m}^2 \text{s}^{-1}$ and $0.4 \mu\text{m}$, respectively, that are very similar to those observed *in vivo*.



Extended Data Figure 3 | Probing the roles of cell envelope systems in the restricted diffusion of OMPs. Laser-scanning confocal FRAP microscopy of *E. coli* cells labelled with CoIE9^{AF488} for BtuB (a) and ColIa^{AF488} or ColIa^{TMR} for Cir (b) was used in conjunction with gene deletions or chemical treatments to probe the involvement of various cell envelope systems in the restricted mobility of OMPs. The rationale for these experiments was twofold. First, structures or processes within the Gram-negative cell envelope could be responsible for the restricted mobility of OMPs (Extended Data Figs 1 and 2). Second, FRAP would provide a means of detecting the loss of restricted OMP mobility, resulting in the recovery of fluorescence after photobleaching, if these systems were perturbed. Each panel in the figure shows DIC microscopy images of the bacterial cell followed by pre-bleach, bleach ($t = 0$) and post-bleach ($t = 3$ min) fluorescent images of the same cell. All microscopy images are $3.03 \times 4.17 \mu\text{m}^2$. Top-to-bottom; CCCP treatment (0.1 mM) of *E. coli* JM83 cells assessed the impact of dissipating the proton-motive force across the inner membrane. *E. coli* BZB1107 cells (*ompR* *vlamB* *ompF*::Tn5) are deficient for the major outer membrane porins OmpF and OmpC. *E. coli* BW25113 *rfaC* (JW3596) is a deep-rough mutant in which the outer core of the LPS is truncated. *E. coli* BW25113 *lpp* (JW1667) is a deletion of Braun's lipoprotein, an outer membrane lipoprotein that is one of the most abundant proteins in *E. coli*. ~40% of Lpp is covalently attached to the underlying peptidoglycan. *E. coli*

BW25113 *tonB* (JW5195) is deleted for TonB, a protein that spans the periplasm and couples the proton-motive force across the inner membrane with transport of nutrients through outer membrane proteins such as BtuB and Cir. ColIa depends on TonB for import into bacteria. *E. coli* BW25113 *tolA* (JW0729) is deleted for TolA, a protein that spans the periplasm and couples proton-motive force with stabilization of the outer membrane. CoIE9 depends on TolA for import into bacteria. Other *E. coli* K-12 deletion strains tested (but not shown) include *pal* (JW0731), an outer membrane lipoprotein, and *ompA* (JW0940), both of which have domains that form non-covalent contacts with the peptidoglycan cell wall. As the data in the figure show, no mutation or condition resulted in the recovery of fluorescence in FRAP experiments from which we infer these systems/processes are not responsible for the restricted mobility of OMPs. Finally, we tested a mutation of CoIE9 in which the first 83 amino acids of the colicin were deleted (Δ^{1-83} CoIE9) but where the inactivating disulfide across the R-domain remained in place. The N-terminal 83 residues contain protein-protein interaction epitopes for OmpF as well as TolB in the periplasm. No change in FRAP behaviour (not shown) was observed demonstrating that interactions made by the colicin at the cell surface are not responsible for the restricted mobility of the OMP to which it is bound. These experiments were done in duplicate (one representative set of images for each condition is presented in the figure).

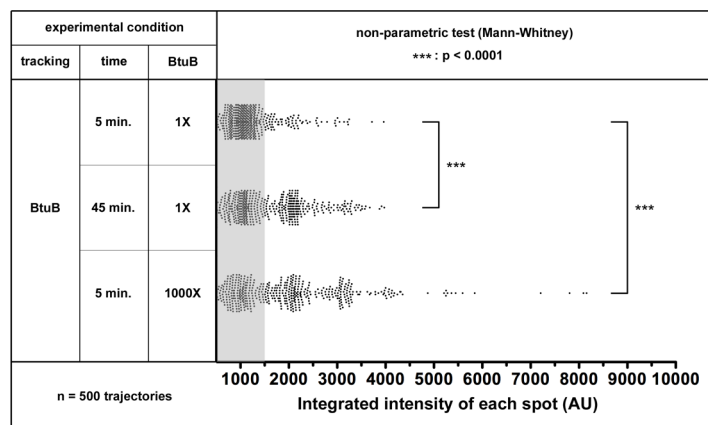


Extended Data Figure 4 | Estimation of the number of BtuB and Cir proteins within OMP islands. **a**, The scatter plots show the distribution of relative intensity of ColE9^{AF488} and ColIa^{TMR} fluorescence signals when bound to their specific OMPs in bacteria or adhered to surfaces. The right-hand panels show representative TIRFM images that were used for the analysis.

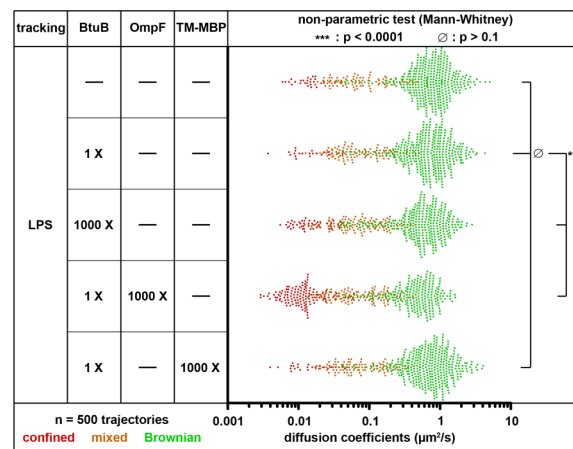
Fluorescence signals corresponded to a stack of 100 frames (30 f.p.s.) and were relatively stable in time and space. Scale bars, 1 μ m. The mean value (\pm s.e.m.) for each distribution is shown by the histograms. We normalized the data for colicins adhered to bacteria using the mean intensity for surface-adhered ColE9^{AF488} and ColIa^{TMR} (300 nM), assuming these correspond to one molecule (\sim 1 AU). From this normalization, OMP islands on average contained \sim 7 labelled OMPs although values ranged from 1 to 19 labelled OMPs. The entire experiment was conducted twice, the replicate showing the same mean value of OMPs per island. **b**, Co-localization of BamA with Cir. TIRFM images (sum of 100 frames) of *E. coli* JWD3 cells expressing HA-tagged

BamA detected by Alexa⁴⁸⁸-labelled anti-HA antibody¹⁹¹⁹. Cir was detected by ColIa^{TMR} labelling before and after a 1-h period of growth. BamA shows significant co-localization ($31 \pm 7\%$ for 112 OMP islands visualized across 20 cells) with Cir within OMP islands. Owing to weak binding of the anti-HA antibody it was not possible to perform more detailed growth experiments as for BamC. **c**, Co-localization of BamC with Cir/TIRFM images (sum of 100 frames) for *E. coli* JM83 cells stained with Alexa⁴⁸⁸-labelled anti-BamC antibody and ColIa^{TMR} showing co-localization within OMP islands that move to the poles in cells undergoing division. Temporal separation of BamC and the Cir labels by a 1-h period of growth, in which old BamC was first blocked with unlabelled antibody, showed that old Cir was localized primarily at the old poles (red label), whereas new BamC-containing islands appeared in non-polar regions of the cell (green label). The average fluorescence distributions and co-localization histograms (error is reported as s.e.m.) shown in **b** and **c** are from 20 cells in each case. * $P < 0.1$ (Student's *t*-test); ** $P < 0.001$ (Mann-Whitney test).

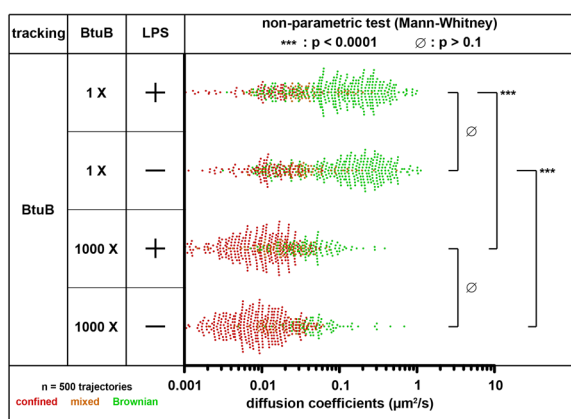
a



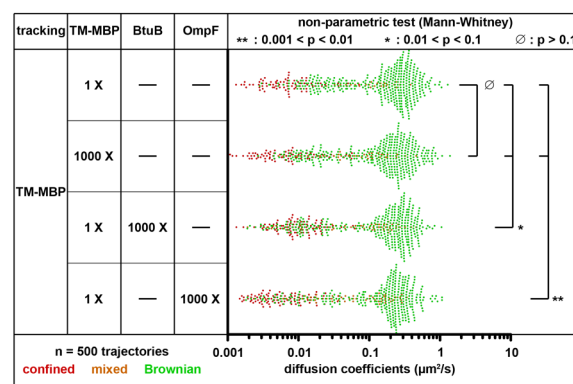
b



c

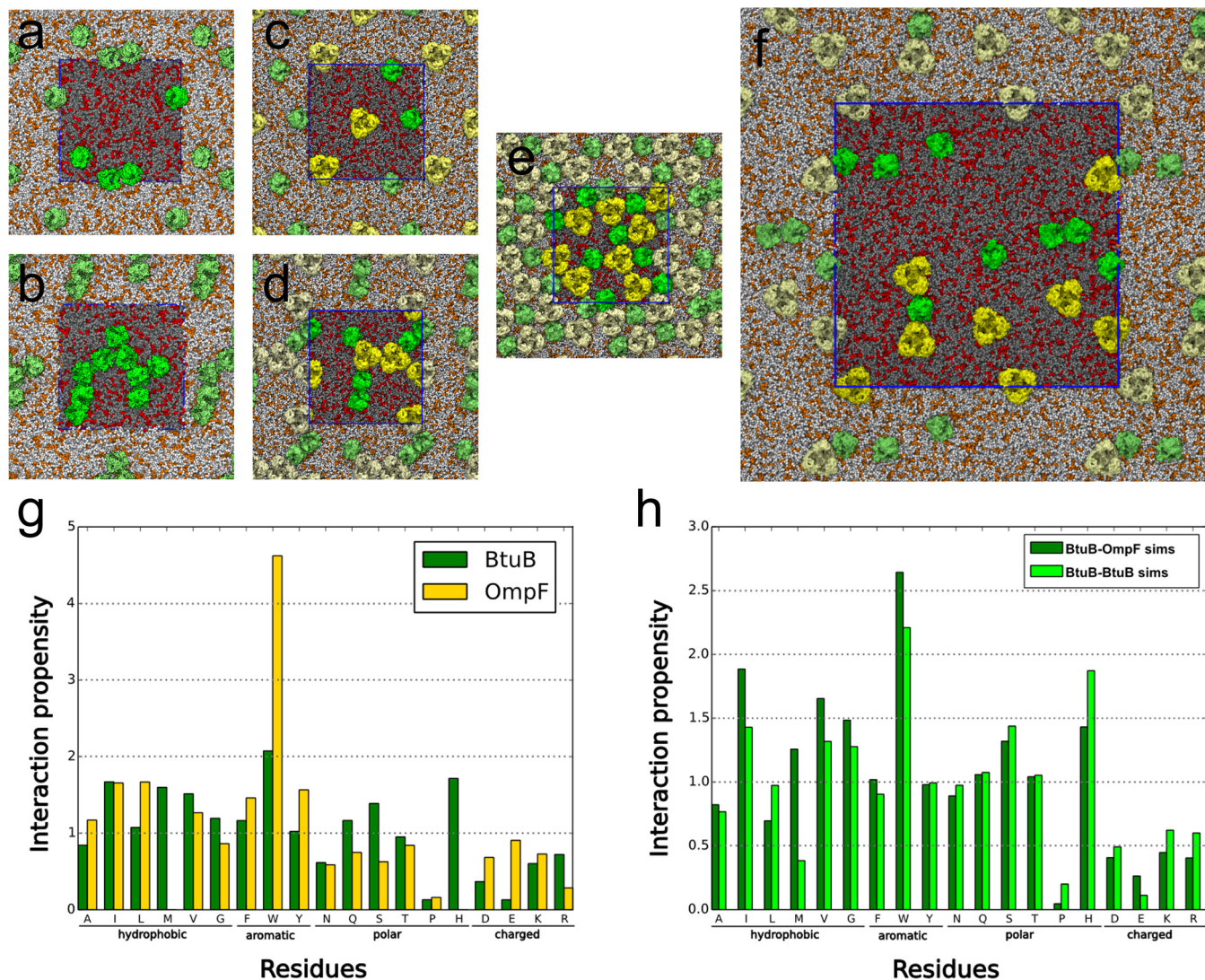


d



Extended Data Figure 5 | Influence of membrane components on diffusion in polymer-supported membranes. **a**, Analysis of the relative fluorescence intensity of BtuB-ColE9^{TMR} suggested significant self-association of BtuB, which was time and concentration dependent. The grey area corresponds to the range of intensities where single step photobleaching was observed and which were the molecules tracked in our experiments. **b**, Distribution of diffusion coefficient when tracking LPS^{BODIPY} incorporated into a polymer-supported bilayer at a ratio of 17:1 ratio LPS:BtuB. High concentrations of BtuB or OmpF, but not of TM-MBP, induced significant trapping of some LPS molecules. **c**, Distribution of diffusion coefficient when tracking BtuB-ColE9^{TMR}

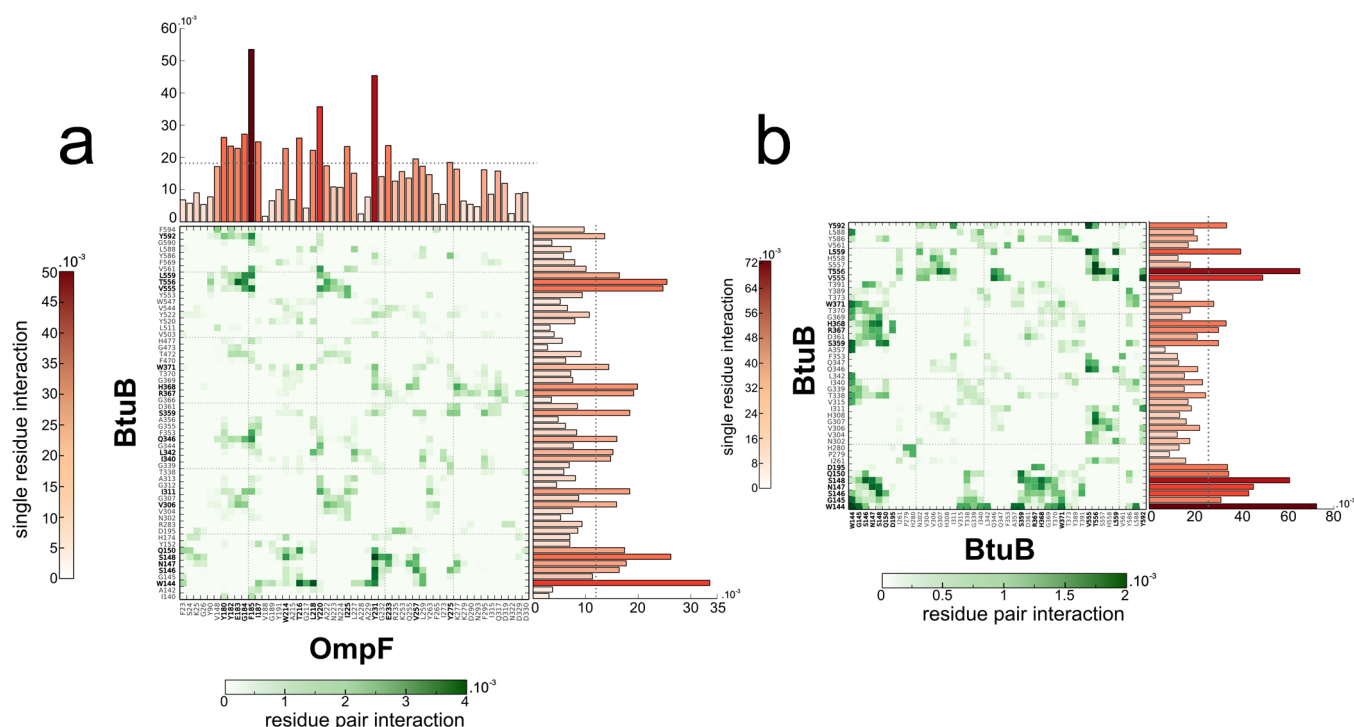
incorporated into a polymer supported bilayer that contains a ratio of 17:1 or 0:1 LPS:BtuB. Addition of LPS did not alter significantly the diffusion behaviour of BtuB in these artificial membranes. **d**, Distribution of diffusion coefficients when tracking TM-MBP^{Cys} incorporated into a polymer-supported bilayer. High concentrations of BtuB or OmpF, but not of TM-MBP, induced a slight but significant trapping of some TM-MBP molecules. All experiments were done in duplicate (one set are presented in the figure) and differences in diffusion coefficient determined using a non-parametric Mann-Whitney test.



Extended Data Figure 6 | Final snapshots of OmpF (yellow) and BtuB (green) positions for coarse grain molecular dynamics simulations.

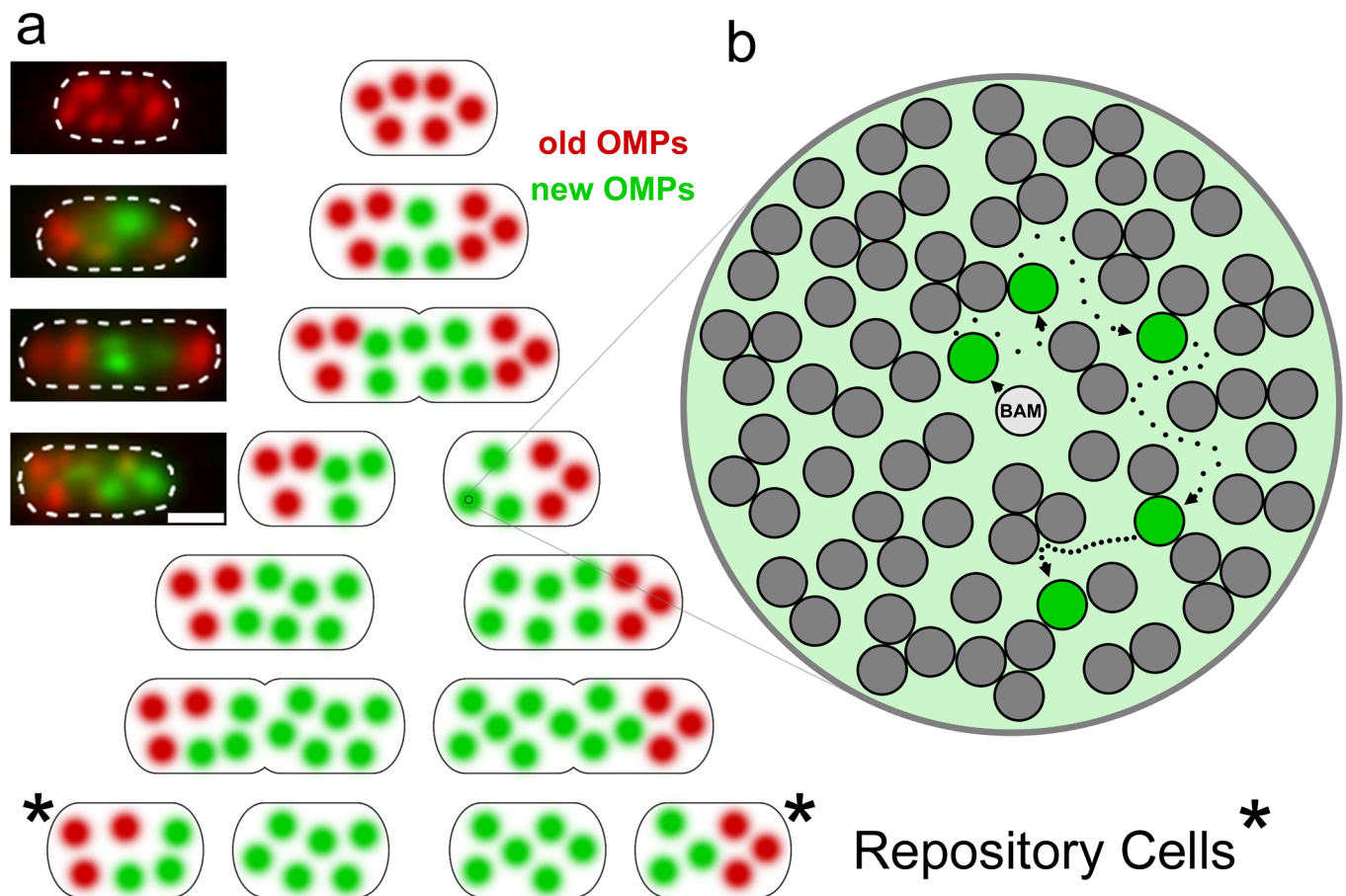
a–f, Each patch ($\sim 30 \times 30 \text{ nm}^2$ for **a–e**, $\sim 60 \times 60 \text{ nm}^2$ for **f**) contained four BtuB (**a**); nine BtuB (**b**); two OmpF trimers and two BtuB monomers (**c**); four OmpF and five BtuB (**d**); eight OmpF and eight BtuB (**e**); and eight OmpF and eight BtuB (**f**). In each case, the single unit cell is darkened and outlined in blue. Lipids are shown in grey (PE) and red (PG). See Extended Data Fig. 7 and Extended Data Table 2 for further details. **g**, Propensity for BtuB (green) and OmpF (yellow) residues to be at the protein–protein interface based on simulations of BtuB–OmpF mixtures. A propensity greater than one indicates

that a residue occurs more frequently at the interaction interface than on the protein surface. **h**, BtuB residue propensities for residue types at the interface between two BtuB monomers, based on simulations containing OmpF and BtuB (dark green), or just BtuB (light green). The two propensities are similar. Note that the sampling in terms of homo-interactions of BtuBs in the OmpF–BtuB simulation is less than in the simulation containing only BtuB owing to the starting positions of the proteins in OmpF–BtuB simulations, which favour hetero-interactions between BtuB and OmpF. See Methods for details about the calculation of the propensity values.



Extended Data Figure 7 | Residues that mediate BtuB–OmpF and BtuB–BtuB interactions in molecular dynamics simulations. **a, b,** Residues that mediate BtuB–OmpF (**a**) and BtuB–BtuB (**b**) interactions. The interaction matrix charts the frequency of interaction between any pair of BtuB and OmpF residues, as a proportion of the total number of interactions that occurred, from high proportional frequency (dark green) to low (white). Depicted here is a subset of the entire interaction matrix, showing only the residues which engaged in interactions with the other protein over a threshold value; any BtuB residue which had a proportional interaction frequency of more than 1×10^{-3}

with any OmpF residue is shown, and similarly for any OmpF residue. On each side, residues with interaction frequency values above approximately one-third of the maximum value of interaction are highlighted in bold. The bar plots show the proportional interaction frequencies of each single BtuB (side) and OmpF (top) residue, for the subset of residues that are shown in the interaction matrix. See Methods for a full mathematical explanation of the interaction value calculations. Bar plots are coloured according to the bar values, from high proportional interaction frequency (dark red) to low (white). This is consistent with the colour scheme in Fig. 4e. Note the matrix in **b** is symmetric.



Extended Data Figure 8 | Binary OMP partitioning and its relationship to biogenesis and organization of OMP islands. **a**, Model depicting the appearance of new OMP islands (green) within growing cells containing old OMP islands (red) in which OMP biogenesis has ceased. The TIRFM images of single cells shown alongside the different stages were taken from cell growth experiments (same experiments as shown in Fig. 2, main text). The model highlights how new OMP islands appear mostly at mid-cell, their creation in conjunction with cell elongation forces old OMP islands towards the poles. The result is binary turnover when the cell divides. The bulk of the old OMPs are retained at the poles of repository cells, which are created at every cell division. It remains unclear why OMP biogenesis ceases in old OMP islands. It is also unclear whether the architecture of the poles (for example, membrane curvature) has a role in retaining old OMPs or whether this is entirely a consequence of continued growth and biogenesis in daughter cells. Nevertheless, the outcome of such a mechanism is that within just two divisions cells appear that do not have any of the original old OMPs. The stochastic nature of the process occasionally results in some old OMP islands not

partitioning with the old pole (for example, main text Fig. 2b, panel 3, TIRFM image), which likely explains why the process is not a pure binary mechanism (if it were, values of 100% would be expected in the histogram shown in Fig. 1f, main text). We speculate that the lack of intermixing between OMP islands, which is consistent with the absence of fluorescence recovery in *in vivo* FRAP experiments (Extended Data Fig. 1), may be due to the very high density of OMPs (particularly porins)²⁵ within islands separated by densely packed LPS. **b**, Model depicting the organization of an individual OMP island and the movement of an OMP within it. OMP islands contain one or more Bam complexes (see main text Fig. 3). Here, we show a single Bam machine having just deposited an OMP (green circle) in the OM. The new OMP diffuses laterally in the membrane but becomes increasingly restricted due to promiscuous interactions with other OMPs (grey circles), which is consistent with the confinement experienced by OMPs *in vitro* and *in vivo* (main text Fig. 4b and Extended Data Fig. 2) and with Monte Carlo and molecular dynamics simulations (main text Fig. 4c, d and Extended Data Figs 2c, 6 and 7). Only a fraction of the OMPs presumed to be present within an OMP island are shown.

Extended Data Table 1 | Comparison of BtuB and Cir diffusion parameters from the present work with those reported for other Gram-negative OMPs and IMPs in the literature

Protein	D ($\mu\text{m}^2/\text{s}$)	Confinement diameter (μm)	Reference
LamB	0.15	0.03	55
BtuB	0.05, 0.10	NR	56
OmpF	0.006	0.1	56
BtuB (ColE9, JM83 cells)	0.013 (N = 62)	0.6	Present work
BtuB (Δ^{1-52} ColE9, JM83 cells)	0.0081 (N = 51)	0.5	Present work
BtuB (ColE9, BZB1107 cells)	0.018 (N = 30)	0.5	Present work
Cir (Colla, JM83 cells)	0.019 (N = 41)	0.6	Present work
Cir (Colla, BZB1107 cells)	0.011 (N = 14)	0.5	Present work
PleC-YFP	0.012		39
MotB-GFP	0.009		40
TatA-YFP	0.1-0.01		57

NR, not reported

OMP and IMP data are shown above and below the grey bar, respectively. Diffusion coefficients (D) were obtained from linear regression of the first 4–5 time steps. See Methods and Extended Data Fig. 2 for details. Two strains were used for the single molecule studies reported in the present paper, the standard lab strain *E. coli* JM83 and a strain in which the major porins OmpF and OmpC had been deleted, *E. coli* BZB1107 (ref. 58). BtuB labelling was achieved using either full-length, fluorescently labelled ColE9 or a variant lacking the first 52 amino acids, a disordered region of the toxin that contains an OmpF binding site and a binding epitope for the periplasmic protein TolB¹⁰. Cir labelling was achieved using fluorescently labelled Colla. The single-molecule trajectories analysed for each condition represent a minimum of four experimental replicates.

Extended Data Table 2 | Summary of coarse grain molecular dynamics simulations for BtuB or OmpF systems and OmpF/BtuB mixtures at differing levels of crowding.

Number of proteins		Dimensions (nm ²)	Protein fractional ratio (%)	Simulation time (μs)	Diffusion coefficient (μm ² /s)	
<i>BtuB</i>	<i>OmpF</i>				<i>BtuB</i>	<i>OmpF</i>
1	0	30 x 30	2	10	11.53 ± 0.33	---
4	0	30 x 30	9	10	8.16 ± 0.78	---
9	0	30 x 30	19	10	2.88 ± 0.94	---
36	0	60 x 60	19	10	6.71 ± 1.00	---
0	1	30 x 30	4	10	---	6.64 ± 1.72
2	2	30 x 30	13	10	9.46 ± 0.78	5.17 ± 1.55
8	8	60 x 60	13	10	13.94 ± 2.64	11.58 ± 1.29
5	4	30 x 30	28	10	2.35 ± 0.06	1.89 ± 0.18
18	18	60 x 60	29	10	2.25 ± 1.74	2.22 ± 1.38
8	8	30 x 30	52	10	0.27 ± 0.00	0.26 ± 0.03

See Methods and Extended Data Figs 6 and 7 for details. Increasing the density of proteins decreased their calculated diffusion coefficients, although the magnitude of the change was contingent on the size of the system, in agreement with recent modelling studies on crowded membranes^{24,54} and with the PSM data shown in Fig. 4b. While diffusion coefficients are qualitatively in agreement with our experimental results the time scale remains too short to enable quantitative comparisons, as has been discussed in a number of other coarse-grained molecular dynamics studies⁵⁴.

Structural and functional features of central nervous system lymphatic vessels

Antoine Louveau^{1,2}, Igor Smirnov^{1,2}, Timothy J. Keyes^{1,2}, Jacob D. Eccles^{3,4,5}, Sherin J. Rouhani^{3,4,6}, J. David Peske^{3,4,6}, Noel C. Derecki^{1,2}, David Castle⁷, James W. Mandell⁸, Kevin S. Lee^{1,2,9}, Tajie H. Harris^{1,2} & Jonathan Kipnis^{1,2,3}

One of the characteristics of the central nervous system is the lack of a classical lymphatic drainage system. Although it is now accepted that the central nervous system undergoes constant immune surveillance that takes place within the meningeal compartment^{1–3}, the mechanisms governing the entrance and exit of immune cells from the central nervous system remain poorly understood^{4–6}. In searching for T-cell gateways into and out of the meninges, we discovered functional lymphatic vessels lining the dural sinuses. These structures express all of the molecular hallmarks of lymphatic endothelial cells, are able to carry both fluid and immune cells from the cerebrospinal fluid, and are connected to the deep cervical lymph nodes. The unique location of these vessels may have impeded their discovery to date, thereby contributing to the long-held concept of the absence of lymphatic vasculature in the central nervous system. The discovery of the central nervous system lymphatic system may call for a reassessment of basic assumptions in neuroimmunology and sheds new light on the aetiology of neuroinflammatory and neurodegenerative diseases associated with immune system dysfunction.

Seeking to identify routes responsible for the recirculation of surveying meningeal immune cells, we investigated the meningeal spaces and the immune cells that occupy those spaces. First, a whole-mount preparation of dissected mouse brain meninges was developed (Fig. 1a) and stained by immunohistochemistry for endothelial cells (Extended Data Fig. 1a), T cells (Fig. 1b) and major histocompatibility complex II (MHCII)-expressing cells (Extended Data Fig. 1b). Labelling of these cells revealed a restricted partitioning of immune cells throughout the meningeal compartments, with a high concentration of cells found in close proximity to the dural sinuses (Fig. 1b; Extended Data Fig. 1b–d).

The dural sinuses drain blood from both the internal and the external veins of the brain into the internal jugular veins. The exact localization of the T lymphocytes around the sinuses was examined to rule out the possibility of artefacts caused by incomplete intracardial perfusion. Coronal sections of the dura mater (Fig. 1c, d) were stained for CD3e (T cells) and for CD31 (endothelial cells). Indeed, the vast majority of the T lymphocytes near the sinuses were abluminal (Fig. 1e). To confirm this finding, mice were injected intravenously (i.v.) with DyLight 488 lectin or fluorescent anti-CD45 antibody before euthanasia and the abluminal localization was confirmed (Extended Data Fig. 1e, f) and quantified (Fig. 1f). Unexpectedly, a portion of T cells (and of MHCII-expressing cells) was aligned linearly in CD31-expressing structures along the sinuses (only few cells were evident in meningeal blood vessels of similar diameter), suggesting a unique function for these perisinusoidal vessels (Fig. 1g–i).

In addition to the cardiovascular system, the lymphatic vessels represent a distinct and prominent vascular system in the body^{7,8}.

Prompted by our observations, the perisinusoidal vessels were tested for markers associated with lymphatic endothelial cells (LEC). Whole-mount meninges from adult mice were immunostained for the LEC marker, Lyve-1. Two to three Lyve-1-expressing vessels were identified running parallel to the dural sinuses (Fig. 1j, k). Analysis of coronal sections labelled for Lyve-1 and the endothelial cell marker, CD31, revealed that Lyve-1 vessels are located adjacent to the sinus (Fig. 1l) and exhibit a distinct lumen (Fig. 1m). Intravenous injection of DyLight 488 lectin before euthanasia confirmed that these Lyve-1⁺ vessels do not belong to the cardiovascular system (Extended Data Fig. 1g, Supplementary Video 1).

The lymphatic character of the perisinusoidal vessels was further interrogated by assessing the presence of several classical LEC markers. Expression of the main LEC transcription factor, Prox1, was indeed detectable in the Lyve-1⁺ vessels using immunostaining in wild-type mice (Extended Data Fig. 2a) or in transgenic mice expressing tdTomato (tdT) under the Prox1 promoter (Prox1^{tdT}; Fig. 2a). Similar to peripheral lymphatic vessels, the Lyve-1 vessels were also found to express podoplanin (Fig. 2b, Extended Data Fig. 2b, c) and the vascular endothelial growth factor receptor 3 (VEGFR3) (Fig. 2c, Extended Data Fig. 2d). Injection of VEGFR3-specific recombinant VEGF-c into the cisterna magna resulted in an increase in the diameter of the meningeal lymphatic vessels, when examined 7 days after the injection (Fig. 2d, e, Extended Data Fig. 2e), suggesting a functional role of VEGFR3 on meningeal LECs. Finally, the presence of LECs in the meninges was confirmed by flow cytometry; a CD45⁺CD31⁺podoplanin⁺ population of cells (LECs) was detected in the dura mater, and is similar to that found in the skin and diaphragm (Extended Data Fig. 3). We identified a potentially similar structure in human dura (Lyve-1⁺podoplanin⁺CD68⁺; Extended Data Fig. 4), but further studies will be necessary to fully assess and characterize the location and organization of meningeal lymphatic vessels in the human central nervous system.

Two types of afferent lymphatic vessels exist—initial and collecting. They differ anatomically (that is, the presence or absence of surrounding smooth muscle cells and lymphatic valves), in their expression pattern of adhesion molecules^{9,10}, and in their permissiveness to fluid and cell entry⁹. In contrast to the sinuses, the meningeal lymphatic vessels are devoid of smooth muscle cells (Fig. 2f, g). Furthermore, meningeal lymphatic vessels were also positive for the immune-cell chemoattractant protein, CCL21 (refs 11, 12; Extended Data Fig. 5a). Unlike the blood vessels that exhibit a continuous pattern of Claudin-5 and vascular endothelial (VE)-cadherin, the meningeal lymphatic vessels exhibit a punctate expression pattern of these molecules similarly to diaphragm lymphatic vessels⁹ (Extended Data Fig. 5b–f). Also, expression of integrin- α 9, which is characteristic of lymphatic valves¹³, was not found on meningeal lymphatic vessels, but was readily detectable in

¹Center for Brain Immunology and Glia, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. ²Department of Neuroscience, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. ³Medical Scientist Training Program, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. ⁴Beirne B. Carter Center for Immunology Research, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. ⁵Department of Medicine (Division of Allergy), School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. ⁶Department of Microbiology, Immunology, and Cancer Biology, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. ⁷Department of Cell Biology, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. ⁸Department of Pathology (Neuropathology), School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA. ⁹Department of Neurosurgery, School of Medicine, University of Virginia, Charlottesville, Virginia 22908, USA.

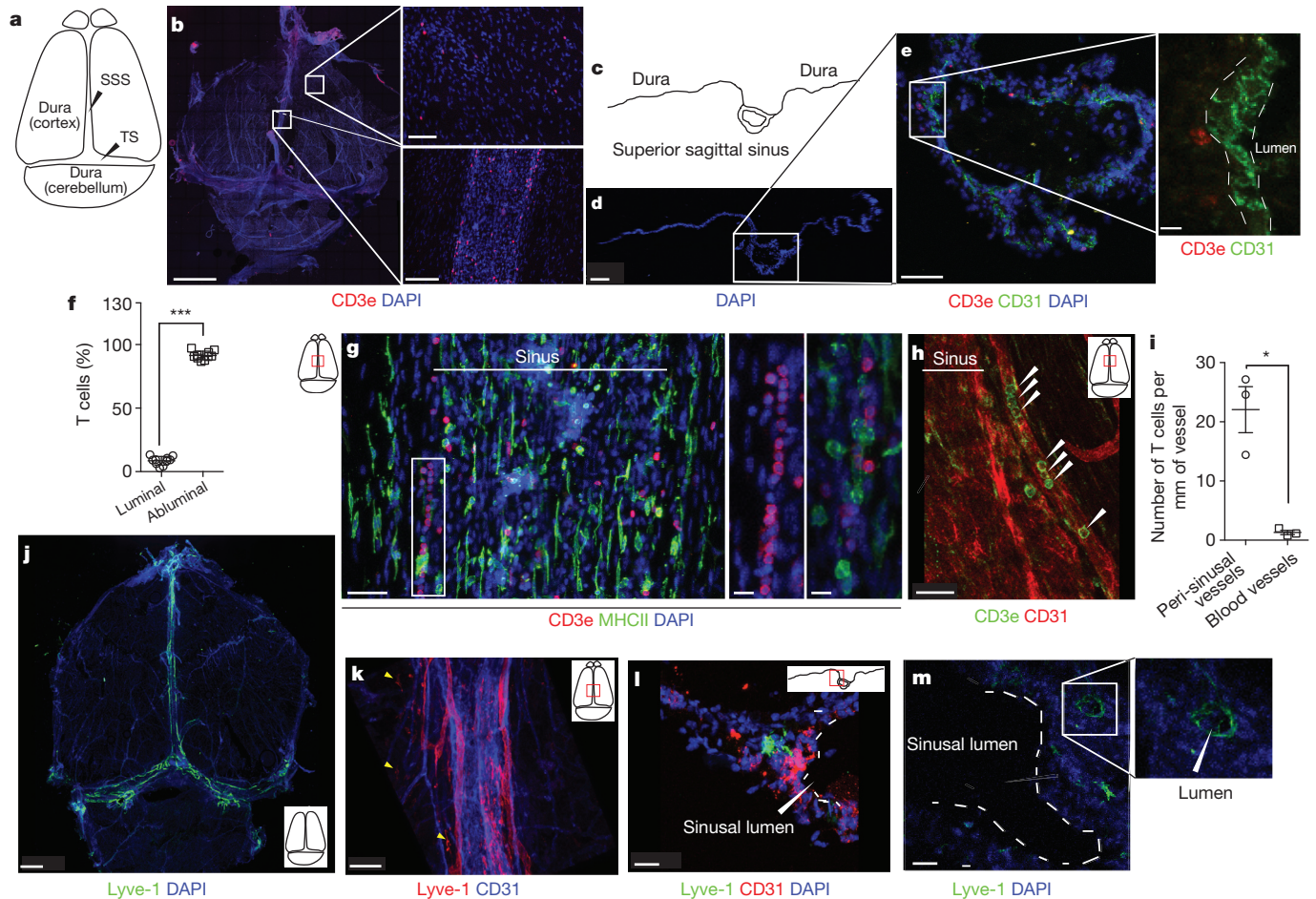


Figure 1 | Abluminal distribution of meningeal T cells and identification of Lyve-1 expressing vessels adjacent to the dural sinuses. **a**, Schematic representation of the whole-mount dissection of the dura mater. SSS, superior sagittal sinus; TS, transverse sinus. **b**, Representative images of CD3e labelling in whole-mount meninges (scale bar, 2,000 µm). Insets, higher magnification of the boxes highlighted in **b** (scale bar, 90 µm (top inset) or 150 µm (bottom inset)). DAPI, 4',6-diamidino-2-phenylindole. **c**, Schematic representation of a coronal section of whole-mount meninges. **d**, Representative image of a coronal section of whole-mount meninges (scale bar, 200 µm). **e**, Representative images of CD3e and CD31 immunolabelling in a coronal section of whole-mount meninges. Scale bar, 100 µm. Inset, higher magnification of the box highlighted in **e** (scale bar, 30 µm, inset, 4 µm). **f**, Quantification of the percentage of sinus T cells localized abuminally vs. luminally to the superior sagittal sinus (mean ± s.e.m.; $n = 18$ fields analysed from 3 independent animals; *** $P = 0.0008$, Mann-Whitney test). **g**, Left, representative images of CD3e and MHCII-expressing cells around the superior sagittal sinus (meningeal

cartoons here and elsewhere depict the location of the presented images; scale bar, 50 µm). Middle, higher magnification of the box highlighted on the left (scale bar, 10 µm). Right, high magnification of CD3- and MHCII-expressing cells (scale bar, 10 µm). **h**, Representative image of CD31 and CD3e labelling around the superior sagittal sinus (scale bar, 30 µm). **i**, Quantification of the number of T cells per mm of vessels in the perisinus CD31⁺ vessels and in similar diameter meningeal blood vessels (mean ± s.e.m.; $n = 3$ animals; * $P = 0.05$, one-tailed Mann-Whitney test). **j**, Representative image of Lyve-1 labelling on whole-mount meninges (scale bar, 1,000 µm). **k**, Higher magnification of Lyve-1-expressing vessels (scale bar, 70 µm); arrowheads indicate Lyve-1⁺ macrophages. **l**, Representative images of CD31 and Lyve-1 labelling of a coronal section of the superior sagittal sinus (scale bar, 70 µm). **m**, Higher magnification of a Lyve-1-positive vessel presenting a conduit-like structure (scale bar, 50 µm). Inset, $\times 1.7$ magnification of the Lyve-1⁺ vessel presented in **m**; arrowhead points to the lumen of the vessel.

the skin lymphatic network (Extended Data Fig. 5g, h). Collectively, these findings indicate that the meningeal lymphatic vessels possess anatomical and molecular features characteristic of initial lymphatic vessels. Furthermore, electron microscopy of whole-mount meninges revealed typical ultrastructural characteristics¹⁴ of the lymphatic vessels, which exhibited a non-continuous basement membrane surrounded by anchoring filaments (Fig. 2h, Extended Data Fig. 5i).

While possessing many of the same attributes as peripheral lymphatic vessels, the general organization and distribution of the meningeal lymphatic vasculature displays certain unique features. The meningeal lymphatic network appears to start from both eyes and track above the olfactory bulb before aligning adjacent to the sinuses (Supplementary Video 2). Compared to the diaphragm, the meningeal lymphatic network covers less of the tissue and forms a less complex network composed of narrower vessels (Extended Data Fig. 5j; Supplementary

Video 2). The vessels are larger and more complex in the transverse sinuses than in the superior sagittal sinus (Extended Data Fig. 5j). The differences in the vessel network could be due to the environment in which the vessels reside—the high cerebrospinal fluid pressure in the central nervous system compared to the interstitial fluid pressure in peripheral tissues could affect the branching of the vessels and also limit their expansion.

Next, the functional capability of the meningeal lymphatic vessels to carry fluid and cells from the meninges/cerebrospinal fluid was examined. Anaesthetized adult mice were simultaneously injected with fluorescein i.v. and with fluorescent tracer dye (QDot655) intracerebroventricularly (i.c.v.), and then imaged through thinned skull by multiphoton microscopy. Vessels filled with QDot655, but not with fluorescein, were seen aligned along the superior sagittal sinus (Fig. 3a–d; Supplementary Video 3), suggesting that non-cardiovascular vessels drain cerebrospinal fluid. This cerebrospinal fluid drainage into

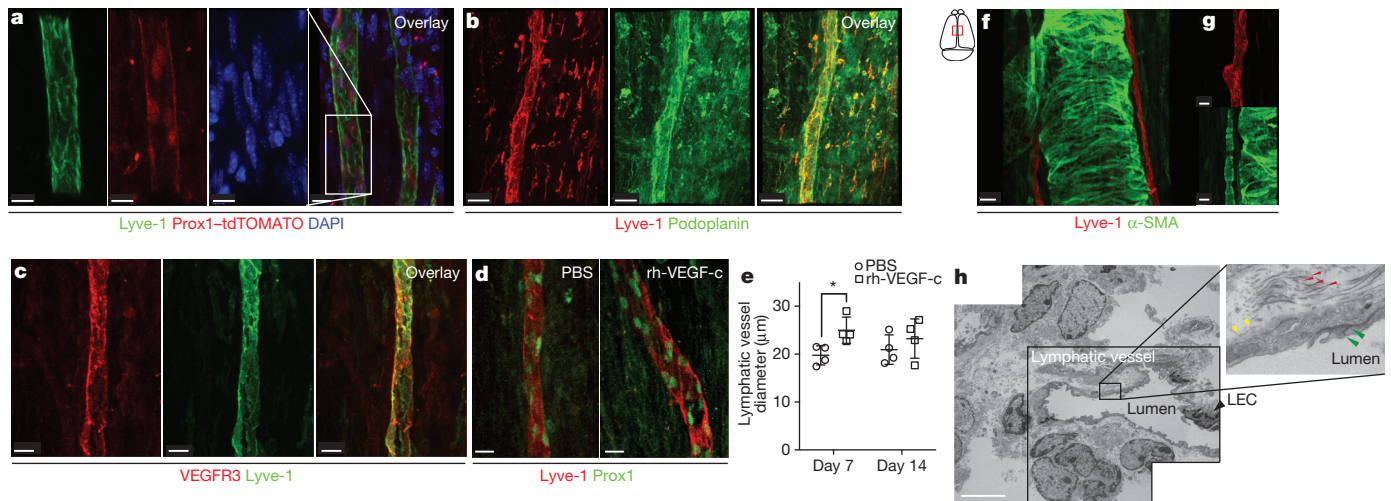


Figure 2 | Molecular and structural characterization of meningeal lymphatic vessels. **a**, Representative images of Prox1 expression in the nuclei of Lyve-1⁺ vessels in the dural sinuses of Prox1^{tdT} mice (scale bars, 10 μ m). **b**, Representative images of podoplanin and Lyve-1 labelling on dural sinuses (scale bar, 40 μ m). **c**, Representative images of VEGFR3 and Lyve-1 staining on dural sinuses (scale bar, 20 μ m). **d**, **e**, Adult mice were injected i.c.v. (cisterna magna) with 4 μ g of rhVEGF-c (Cys156Ser) or with PBS. Meninges were harvested 7 and 14 days after the injection. **d**, Representative images of Lyve-1 and Prox1 labelling of meninges at day 7 after injection (scale bars, 30 μ m).

e, Quantification of the meningeal lymphatic vessel diameter (mean \pm s.e.m.; $n = 4$ mice each group; $*P < 0.05$, two-way ANOVA with Bonferroni post hoc test). **f**, **g**, Representative images of smooth muscle cells (alpha-smooth muscle actin, α -SMA) and Lyve-1 labelling on dural sinuses (scale bars, 50 μ m (**f**) or 20 μ m (**g**)). **h**, Representative low power micrograph (transmission electron microscopy) of a meningeal lymphatic vessel (scale bar, 5 μ m). Inset, higher magnification of the box highlighted in **h**. Yellow arrowheads, basement membrane; red arrowheads, anchoring filaments (collagen fibres); green arrowheads, cellular junction.

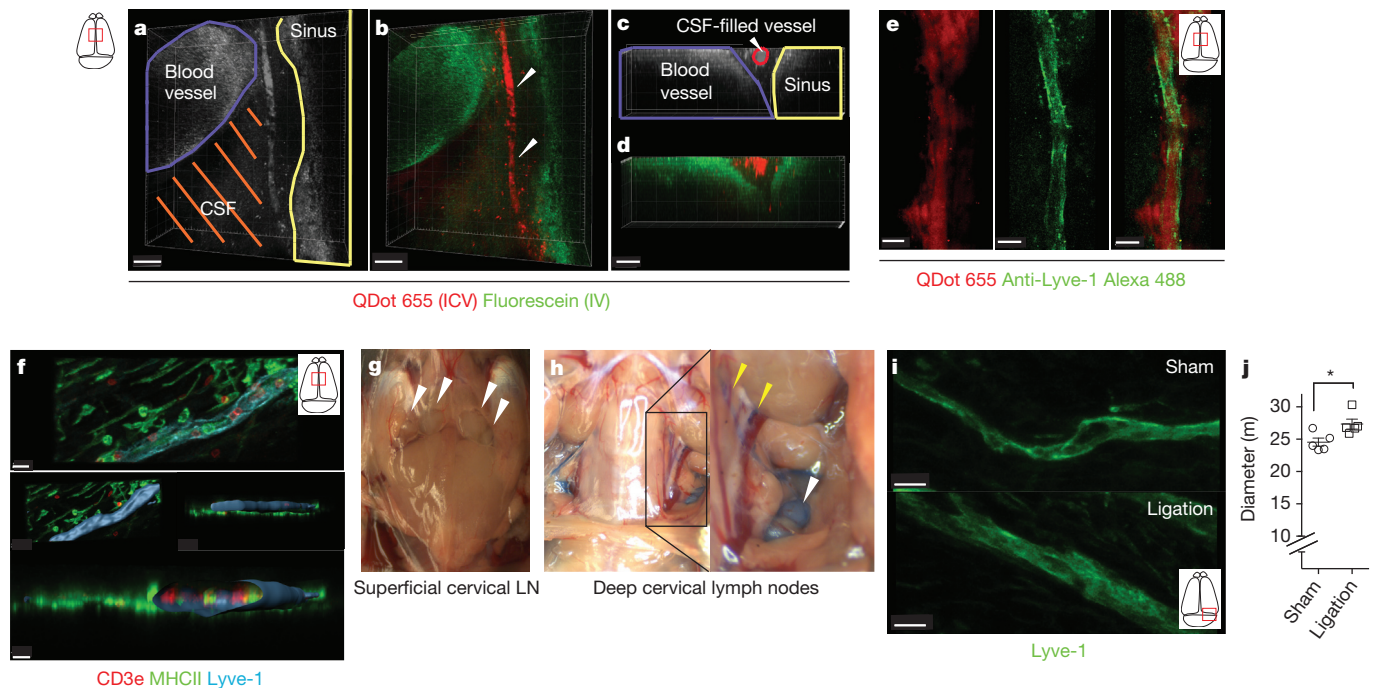


Figure 3 | Functional characterization of meningeal lymphatic vessels. **a**–**d**, Representative z-stacks of the superior sagittal sinus of adult mice injected intravenously (i.v.) with fluorescein and intracerebroventricularly (i.c.v.) with QDot655 ($n = 3$ mice). **a**, **b**, Low-magnification images showing fluorescein labelling in a meningeal blood vessel and in the superior sagittal sinus (scale bars, 20 μ m in **a**, **b**). In contrast, QDot655 labelling is prominent in the perisinusoid vessel. **c**, **d**, Coronal section of the z-stack presented in **a** and **b** (scale bars, 20 μ m in **c**, **d**). **e**, CSF, cerebrospinal fluid. **e**, Representative z-stack of cerebrospinal fluid-filled vessel from a mouse injected i.c.v. with both QDot655 and Alexa488-conjugated anti-Lyve-1 antibody ($n = 3$ mice; scale bars, 30 μ m). **f**, Top, representative image of immunolabelling for CD3e and MHCII along with Lyve-1 in the meninges (scale bar, 15 μ m). Bottom, representative image of a 3D reconstruction of the meningeal lymphatic vessels showing the luminal

localization of the CD3e and MHCII-expressing cells (scale bar, 20 μ m). **g**, **h**, Adult mice were injected i.c.v. with 5 μ l of 10% Evans blue. Superficial cervical lymph nodes (**g**) and deep cervical lymph nodes (**h**) were analysed 30 min after injection ($n = 5$ mice); white arrowheads indicate the lymph nodes; yellow arrowheads indicate the Evans blue-filled vessels arising near the internal jugular vein into the deep cervical lymph nodes (**h**). **i**, **j**, The collecting vessels draining into the deep cervical lymph nodes (yellow arrowheads in **h**) were ligated or sham-operated. Eight hours after the ligation, the meninges were collected and immunolabelled for Lyve-1. Representative images of immunolabelling for Lyve-1 in the transverse sinus of ligated and sham-operated mice (**i**; scale bars, 30 μ m). Dot plots represent measurement of the meningeal lymphatic vessel diameters (**j**; mean \pm s.e.m.; $n = 5$ mice each group from 2 independent experiments; $*P = 0.031$, Mann-Whitney test).

meningeal vessels may occur in addition to the previously described cerebrospinal fluid filtration into the dural sinuses via arachnoid granulations^{15,16} (Extended Data Fig. 6a). Injection of Alexa488-conjugated anti-Lyve-1 antibody i.c.v. labelled the meningeal lymphatic vessels (Extended Data Fig. 6b, c; Supplementary Video 4). Moreover, co-injection of a QDot655 and an Alexa488-conjugated anti-Lyve-1 antibody i.c.v. demonstrated that the meningeal lymphatic vessels were indeed filled with QDot655, and thus were draining the cerebrospinal fluid (Fig. 3e). Imaging of the QDot655-filled lymphatic vessels revealed a slower flow rate but similar direction of flow in the meningeal lymphatic vessels compared to the adjacent blood vessels (Supplementary Video 5), similar to what is observed outside of the central nervous system¹⁷.

Classic lymphatic vessels, in addition to draining interstitial fluids, allow cells to travel from tissues to draining lymph nodes¹⁸. We therefore examined whether the meningeal lymphatic vessels were capable of carrying leukocytes. Immunohistochemical analysis of whole-mount meninges revealed that ~24% of all sinusoidal T cells and ~12% of all sinusoidal MHCII⁺ cells were found within these vessels (Fig. 3f, Extended Data Fig. 7a, b). Moreover, CD11c⁺ cells and B220⁺ cells are also found in the meningeal lymphatic vessels of naive mice (Extended Data Fig. 7c–f).

Cellular and soluble constituents of the cerebrospinal fluid have been shown to elicit immune responses in the cervical lymph nodes^{19–22}. Their proposed path is via the cribriform plate into lymphatic vessels within the nasal mucosa²³. To determine whether meningeal lymphatic vessels communicate with deep cervical lymph nodes directly, we injected mice with Evans blue i.c.v. and examined peripheral lymph nodes for the presence of the dye over a 2 h period. Thirty minutes after injection, Evans blue was detected in the meningeal lymphatic vessels and the sinus, as expected (Extended Data Fig. 8a–c), and had also drained into the deep cervical lymph nodes (dCLN) (Fig. 3g, h), but not into the superficial cervical lymph nodes. At later time points, Evans blue was also present in the superficial cervical lymph nodes (data not shown). Virtually no Evans blue was seen in the surrounding non-lymphatic tissue at the time points tested. Interestingly, no Evans blue was detected in the dCLN 30 min after direct injection into the nasal mucosa (Extended Data Fig. 8d, e), suggesting that meningeal lymphatic vessels and not nasal mucosa lymphatic vessels represent the primary route for drainage of cerebrospinal fluid-derived soluble and cellular constituents into the dCLN during this time frame.

Resection of the dCLN affected the T-cell compartment of the meninges, resulting in an increase in the number of meningeal T cells (Extended Data Fig. 9a–e). This presumably resulted from an inability of T cells to drain from the meningeal spaces, consistent with a direct connection between the meninges and the dCLN. To further demonstrate this connection, we ligated the lymph vessels that drain into the dCLNs (Extended Data Fig. 9f), and injected the mice i.c.v. with Evans blue. No accumulation of Evans blue was evident in the dCLN of ligated mice, as opposed to dye accumulation in the sham-operated controls (Extended Data Fig. 9g). Moreover, an increase in the diameter of the meningeal lymphatic vessels was observed (Fig. 3i, j; Extended Data Fig. 9h), similar to lymphoedema observed in peripheral tissues²⁴. These results further suggest physical connection between the meningeal lymphatic vessels and the dCLN.

Drainage of the cerebrospinal fluid into the periphery has been a subject of interest for decades and several routes have been described regarding how cerebrospinal fluid can leave the central nervous system^{23,25}. The newly discovered meningeal lymphatic vessels are a novel path for cerebrospinal fluid drainage and represent a more conventional path for immune cells to egress the central nervous system. Our findings may represent the second step in the drainage of the interstitial fluid from the brain parenchyma into the periphery after it has been drained into the

cerebrospinal fluid through the recently discovered glymphatic system^{26,27} (Extended Data Fig. 10).

The presence of a functional and classical lymphatic system in the central nervous system suggests that current dogmas regarding brain tolerance and the immune privilege of the brain should be revisited. Malfunction of the meningeal lymphatic vessels could be a root cause of a variety of neurological disorders in which altered immunity is a fundamental player such as multiple sclerosis, Alzheimer's disease, and some forms of primary lymphoedema that are associated with neurological disorders^{28–30}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 October 2014; accepted 20 March 2015.

Published online 1 June 2015.

1. Ransohoff, R. M. & Engelhardt, B. The anatomical and cellular basis of immune surveillance in the central nervous system. *Nature Rev. Immunol.* **12**, 623–635 (2012).
2. Kipnis, J., Gadani, S. & Derecki, N. C. Pro-cognitive properties of T cells. *Nature Rev. Immunol.* **12**, 663–669 (2012).
3. Shechter, R., London, A. & Schwartz, M. Orchestrated leukocyte recruitment to immune-privileged sites: absolute barriers versus educational gates. *Nature Rev. Immunol.* **13**, 206–218 (2013).
4. Goldmann, J. et al. T cells traffic from brain to cervical lymph nodes via the cribriform plate and the nasal mucosa. *J. Leukoc. Biol.* **80**, 797–801 (2006).
5. Kaminski, M. et al. Migration of monocytes after intracerebral injection at entorhinal cortex lesion site. *J. Leukoc. Biol.* **92**, 31–39 (2012).
6. Engelhardt, B. & Ransohoff, R. M. The ins and outs of T-lymphocyte trafficking to the CNS: anatomical sites and molecular mechanisms. *Trends Immunol.* **26**, 485–495 (2005).
7. Alitalo, K. The lymphatic vasculature in disease. *Nature Med.* **17**, 1371–1380 (2011).
8. Wang, Y. & Oliver, G. Current views on the function of the lymphatic vasculature in health and disease. *Genes Dev.* **24**, 2115–2126 (2010).
9. Baluk, P. et al. Functionally specialized junctions between endothelial cells of lymphatic vessels. *J. Exp. Med.* **204**, 2349–2362 (2007).
10. Kerjaschki, D. The lymphatic vasculature revisited. *J. Clin. Invest.* **124**, 874–877 (2014).
11. Debes, G. F. et al. Chemokine receptor CCR7 required for T lymphocyte exit from peripheral tissues. *Nature Immunol.* **6**, 889–894 (2005).
12. Weber, M. et al. Interstitial dendritic cell guidance by haptotactic chemokine gradients. *Science* **339**, 328–332 (2013).
13. Bazigou, E. et al. Integrin- $\alpha 9$ is required for fibronectin matrix assembly during lymphatic valve morphogenesis. *Dev. Cell* **17**, 175–186 (2009).
14. Koina, M. E. et al. Evidence for lymphatics in the developing and adult human choroid. *Invest. Ophthalmol. Vis. Sci.* **56**, 1310–1327 (2015).
15. Johanson, C. E. et al. Multiplicity of cerebrospinal fluid functions: new challenges in health and disease. *Cerebrospinal Fluid Res.* **5**, 10 (2008).
16. Weller, R. O., Djuanda, E., Yow, H.-Y. & Carare, R. O. Lymphatic drainage of the brain and the pathophysiology of neurological disease. *Acta Neuropathol.* **117**, 1–14 (2009).
17. Liu, N.-F., Lu, Q., Jiang, Z.-H., Wang, C.-G. & Zhou, J.-G. Anatomic and functional evaluation of the lymphatics and lymph nodes in diagnosis of lymphatic circulation disorders with contrast magnetic resonance lymphangiography. *J. Vasc. Surg.* **49**, 980–987 (2009).
18. Girard, J.-P., Moussion, C. & Förster, R. HEVs, lymphatics and homeostatic immune cell trafficking in lymph nodes. *Nature Rev. Immunol.* **12**, 762–773 (2012).
19. Weller, R. O., Galea, I., Carare, R. O. & Minagar, A. Pathophysiology of the lymphatic drainage of the central nervous system: implications for pathogenesis and therapy of multiple sclerosis. *Pathophysiology* **17**, 295–306 (2010).
20. Mathieu, E., Gupta, N., Macdonald, R. L., Ai, J. & Yücel, Y. H. In vivo imaging of lymphatic drainage of cerebrospinal fluid in mouse. *Fluids Barriers CNS* **10**, 35 (2013).
21. Cserr, H. F., Harling-Berg, C. J. & Knopf, P. M. Drainage of brain extracellular fluid into blood and deep cervical lymph and its immunological significance. *Brain Pathol.* **2**, 269–276 (1992).
22. Harris, M. G. et al. Immune privilege of the CNS is not the consequence of limited antigen sampling. *Sci. Rep.* **4**, 4422 (2014).
23. Laman, J. D. & Weller, R. O. Drainage of cells and soluble antigen from the CNS to regional lymph nodes. *J. Neuroimmune Pharmacol.* **8**, 840–856 (2013).
24. Schneider, M., Ny, A., Ruiz de Almodovar, C. & Carmeliet, P. A new mouse model to study acquired lymphoedema. *PLoS Med.* **3**, e264 (2006).
25. Kida, S., Pantazis, A. & Weller, R. O. CSF drains directly from the subarachnoid space into nasal lymphatics in the rat. Anatomy, histology and immunological significance. *Neuropathol. Appl. Neurobiol.* **19**, 480–488 (1993).
26. Xie, L. et al. Sleep drives metabolite clearance from the adult brain. *Science* **342**, 373–377 (2013).
27. Yang, L. et al. Evaluating glymphatic pathway function utilizing clinically relevant intrathecal infusion of CSF tracer. *J. Transl. Med.* **11**, 107 (2013).

28. Berton, M. *et al.* Generalized lymphedema associated with neurologic signs (GLANS) syndrome: A new entity? *J. Am. Acad. Dermatol.* **72**, 333–339 (2015).
29. Akiyama, H. *et al.* Inflammation and Alzheimer's disease. *Neurobiol. Aging* **21**, 383–421 (2000).
30. Hohlfeld, R. & Wekerle, H. Immunological update on multiple sclerosis. *Curr. Opin. Neurol.* **14**, 299–304 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Smith for editing the manuscript; V. Engelhard for initial discussions and suggestions on lymphatic endothelial cell specific markers. We also thank the members of the Kipnis laboratory, Brain Immunology and Glia (BIG) center, and the Department of Neuroscience at the University of Virginia (especially J. Lukens) for their valuable comments during multiple discussions of this work. L. Holland and B. Lopes provided the human samples. This work was funded by "Fondation pour la Recherche Medicale" to A.L. and by The National Institutes of Health (R01AG034113 and R01NS061973) to J.K.

Author Contributions A.L. performed most of the experiments, analysed the data, and contributed to experimental design and manuscript writing. I.S. performed all the surgeries and intracerebroventricular injections. T.J.K. assisted with the experiments and the analysis of the data. J.D.E, S.J.R. and J.D.P. participated in the discussions and helped with the experimental design. N.C.D. performed the xDCLN experiment. D.C. contributed to the imaging and the analysis of the electron microscopy images. J.W.M. contributed with data analysis of the human samples. K.S.L. contributed to experimental design and to manuscript editing. T.H.H. assisted to the intravital imaging experiment, contributed to experimental design and to manuscript editing. J.K. designed the study, assisted with data analysis, and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.L. (al2hk@virginia.edu) or J.K. (kipnis@virginia.edu).

METHODS

No statistical methods were used to predetermine sample size.

Animals. Male and female C57Bl/6, *NOD.Cd11c-YFP* and *Prox1^{tdT}* mice were purchased from Jackson Laboratories and were housed in temperature and humidity controlled rooms, maintained on a 12 h/12 h light/dark cycle (lights on at 7:00). All strains were kept in identical housing conditions. All procedures complied with regulations of the Institutional Animal Care and Use Committee at University of Virginia. Only adult animals (eight to ten weeks) were used in this study. Sample size was chosen in accordance with similar, previously published experiments. Animals from different cages in the same experimental group were selected to assure randomization. For all experiments, investigators were blinded from the time of euthanasia to the end of the analysis.

Human samples. Autopsy specimens of human dura including the superior sagittal sinus were obtained from the Departments of Pathology and Neurosurgery at the University of Virginia. All samples are from consenting patients that gave no restriction to the use of their body for research and teaching (through the Virginia Anatomical Board, Richmond, VA). All obtained samples were fixed and stored in a 10% formalin solution for prolonged time periods.

Meninges immunohistochemistry. Mice were euthanized with an intraperitoneal (i.p.) injection of Euthasol and perfused with 0.1 M of PBS for 5 min. Skin was removed from the head and the muscle was stripped from the bone. After removal of the mandibles and the skull rostral to maxillae, the top of the skull was removed with surgical scissors. Whole-mount meninges were fixed while still attached to the skull cap in PBS with 2% paraformaldehyde (PFA) for 24 h at 4 °C or in 1:1 ethanol:acetone solution for 20 min at −20 °C depending on the antibody. The dura/arachnoid was then dissected from the skullcap. For analysis of the pia mater, brains extracted from the skull were flash frozen, and 40 µm-thick transverse sections were sliced using a cryostat (Leica). The choroid plexus was dissected out of the ventricles of non-fixed brain and fixed with 2% PFA in PBS for 24 h. For coronal sectioning of whole-mount meninges, 100 µl of Matrigel (Corning) was injected into the dural sinuses before dissection. The meninges (dura mater/arachnoid) were dissected from the skullcap, flash frozen, and 10-µm thick sections were cut using a cryostat, and the slices were placed on gelatin-coated slides.

Whole mounts and sections were incubated with PBS containing 2% of normal serum (either goat or chicken), 1% BSA, 0.1% Triton-X-100 and 0.05% of Tween 20 for 1 h at room temperature, followed by incubation with appropriate dilutions of primary antibodies: anti-CD31 (eBioscience, clone 390, 1:100); anti-CD3e (eBioscience, clone 17A2, 1:500), anti-MHC II (eBioscience, clone M5/114.15.2, 1:500), anti-Lyve-1 (eBioscience, clone ALY7, 1:200), anti-Prox1 (Angiobio, 11-002, 1:500), anti-podoplanin (eBioscience, clone 8.1.1, 1:100), anti-VEGFR3 (R&D Systems, AF743, 1:100), anti-α-SMA (Sigma-Aldrich, clone 1A4, 1:500), anti-VE-Cadherin (eBioscience, clone BV13 1:100), anti-Claudin-5 (Molecular Probes, 352588, 1:200), anti-CCL21 (R&D Systems, AF457, 1:100), anti integrin-α9 (R&D Systems, AF3827, 1:100), anti-B220 (eBioscience, clone RA3-6B2, 1:100), anti CD11c (eBioscience, clone N418, 1:100) overnight at 4 °C in PBS containing 1% BSA and 0.5% Triton-X-100. Whole mounts and sections were then washed 3 times for 5 min at room temperature in PBS followed by incubation with Alexafluor 488/594/647 chicken/goat anti rabbit/goat IgG antibodies (Invitrogen, 1:1,000) or PE/Cy3 conjugated streptavidin (eBioscience, 1:1,000) for 1 h at room temperature in PBS with 1% BSA and 0.5% Triton-X-100. After 5 min in 1:10,000 DAPI reagent, whole mount and section were washed with PBS and mounted with Aqua-Mount (Lerner) under coverslips. For the pre-absorption experiments, the anti-VEGFR3 and anti-podoplanin antibodies were incubated respectively with recombinant mouse VEGFR3 (743-R3-100, R&D Systems) or recombinant mouse podoplanin (3244-PL-050, R&D Systems) at the concentration ratio of 1:10 overnight at 4 °C in PBS containing 1% BSA and 0.5% Triton-X-100 before staining.

For human samples, formalin-fixed superior sagittal sinuses were dissected in 2-mm thick sections and fixed overnight in PBS with 4% PFA. Tissues were then embedded in OCT and 10-µm thick sections were sliced onto gelatin-coated slides on a cryostat (Leica). Antigen retrieval was performed by incubation of the slides for 20 min in sodium citrate buffer pH 6.0 at 80 °C. After washing with PBS, endogenous biotins were blocked with a 30 min incubation in PBS with 3% H₂O₂, then the slides were blocked in PBS containing 2% of normal goat serum, 1% BSA, 0.1% Triton-X-100 and 0.05% Tween 20 for 1 h at room temperature. Slides were then incubated overnight at 4 °C with anti-Lyve-1 (ab36993, Abcam, 1:200), anti-podoplanin (HPA007534, Sigma-Aldrich, 1:200) or anti-CD68 (HPA048982, Sigma-Aldrich, 1:1,000) diluted in PBS with 1% BSA and 0.5% Triton-X-100. Sections were then washed 3 times for 5 min at room temperature in PBS followed by incubation with biotin-conjugated goat anti-rabbit antibody (Jackson ImmunoResearch, 1:1,000) for 1 h at room temperature, then for 30 min at room temperature with ABC kit solutions (Vectastain, Vector Labs). Slides were then incubated with the peroxidase substrate DAB (Sigma-Aldrich) for several

minutes, counterstained with haematoxylin, dehydrated, and mounted in Cytoseal 60 (Thermo Scientific) under coverslips. Nine human samples were labelled and analysed; the lymphatic structures were identified in two of them.

Image analysis. Images were acquired with a Leica TCS SP8 confocal system (Leica Microsystems) using the LAS AF Software. For the images of the complete whole mount, images were acquired with a ×10 objective with 0.25 NA. Other confocal images were acquired using a ×20 objective with 0.70 NA or a ×40 oil immersion objective with 1.30 NA. All images were acquired with a 512 × 512 pixel resolution and with a z-step of 4 µm. Quantitative assessments were performed using FIJI software (NIH). Percentage of luminal T cells was determined by counting the number of T cells with luminal localization in the sinuses area. T cell density was established by dividing the number of T lymphocytes by the area of meninges. Prox1-positive cell density was defined by dividing the number of Prox1⁺ nuclei by the area of lymphatic vessels. Statistical analyses were performed using GraphPad Prism software. Specific statistical tests are presented in the text for each experiment. Outlier samples were eliminated using the Grubbs' test with a significance level of 0.01 (only for the rh-VEGF-c experiment). No estimate of variation between groups was performed.

Electron microscopy. Meninges were harvested as previously described and fixed in 2.5% glutaraldehyde, 2% paraformaldehyde in 0.1 M sodium cacodylate buffer, pH 7.4, and post-fixed in 2% osmium tetroxide in 0.1 M cacodylate buffer with 0.15% potassium ferrocyanide. After rinsing in buffer, the tissue was dehydrated through a series of graded ethanol to propylene oxide, infiltrated and embedded in epoxy resin and polymerized at 70 °C overnight. Semi-thin sections (0.5 µm) were stained with toluidine blue for light microscope examination. Ultrathin sections (80 nm) were cut (Integrated microscopy centre, University of Memphis) and imaged using the Tecnai TF20 TEM with an AMT XR41 camera (Extended Data Fig. 5i; Integrated microscopy centre, University of Memphis) or using the Tecnai F20 TEM with an UltraScan CCD camera (Fig. 2h; Advanced Microscopy core, University of Virginia).

Multiphoton microscopy. Mice were anaesthetized by ketamine/xylazine injection i.p. and injected i.c.v. (into the cisterna magna) with 5 µl of QDot655 (Invitrogen) or 5 µl of Alexa488-conjugated anti-Lyve-1 antibody (ALY7, eBioscience). Mice were co-injected i.v. with 25 µl of 10% fluorescein sodium salt (Sigma-Aldrich) or 5 µl of QDot655 (Invitrogen). Meningeal lymphatic vessels were imaged through a thinned skull preparation. The core temperature of the mice was monitored and maintained at 37 °C. Imaging was performed with a Leica TCS SP8 multiphoton microscopy system (Leica Microsystems) equipped with a Chameleon Ultra II tunable Ti:Sapphire laser (Coherent). GFP and QDot655 were excited with an excitation wavelength of 880 nm. Images were obtained using a ×25 water immersion objective with 0.95 NA and external HyD non-descanned detectors (Leica Microsystems). Four-dimensional imaging data were collected by obtaining images from the x, y, and z-planes over time. The resulting images were analysed with Imaris software (Bitplane).

Labelling of the vascular compartment. To assess the abluminal localization of the sinusoidal T cells, mice were injected i.v. with 10 µg of FITC-conjugated anti-CD45 antibody (eBioscience, clone 30-F11) or control isotype 1 h before euthanasia. To assess that the meningeal lymphatic vessels are not part of the cardiovascular system, mice were injected with 100 µl of DyLight 488 Lycopersicon Esculentum Lectin (Vector Laboratories) 5 min before euthanasia.

In vivo VEGFR3 activation. Mice were injected i.c.v. with 4 µg of rh-VEGF-c (Cys156Ser, R&D Systems) or with PBS. Meninges were harvested 7 and 14 days after the injection.

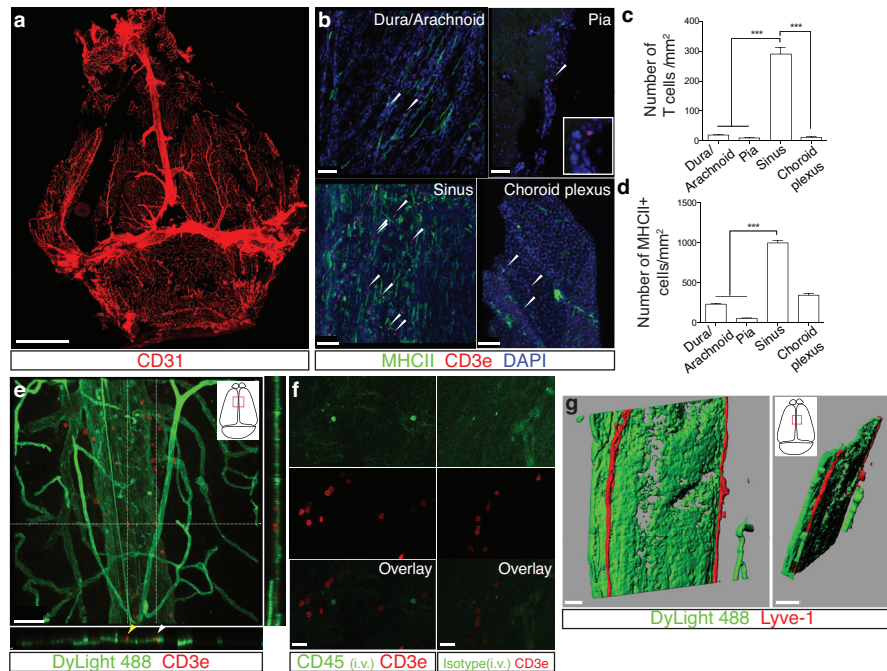
Evans blue injection and detection. Mice were anaesthetized by ketamine/xylazine injection i.p., and then 5 µl of 10% Evans blue (Sigma-Aldrich) was injected i.c.v. into the cisterna magna or intranasally. Thirty minutes after injection, mice were euthanized with CO₂ and the central nervous system draining lymph nodes were dissected for assessment of Evans blue content. The dura mater was also harvested to analyse Evans Blue localization using confocal microscopy. The intensity of the Evans blue was measured using the plot profile function of FIJI.

Flow cytometric analysis of meninges. Mice were perfused with 0.1 M PBS for 5 min. Heads were removed and skulls were quickly stripped. Mandibles were removed, as well as all skull material rostral to maxillae. Surgical scissors were used to remove the top of the skull, cutting clockwise, beginning and ending inferior to the right post-tympanic hook. Meninges (dura mater, arachnoid and pia mater) were carefully removed from the interior aspect of the skulls and surfaces of the brain with Dumont #5 forceps (Fine Science Tools). Meninges were gently pressed through 70 µm nylon mesh cell strainers with sterile plastic plunger (BD Biosciences) to yield a single-cell suspension. For lymphatic endothelial cells isolation, meninges (along with diaphragm and ear skin) were digested for 1 h in 0.41 U ml^{−1} of Liberase TM (Roche) and 60 U ml^{−1} of DNase I. Cells were then centrifuged at 280g at 4 °C for 10 min, the supernatant was removed and cells were resuspended in ice-cold FACS buffer (pH 7.4; 0.1 M PBS; 1 mM EDTA; 1% BSA). Cells were stained for extracellular marker with antibodies to CD45-PacificBlue

(BD Bioscience), CD45-phycoerythrin (PE)-Cy7 or eFluor 450 (eBioscience), TCR β -Alexa780 (eBioscience), CD4-Alexa488 (eBioscience), CD8-peridinin chlorophyll (PerCP)-Cy5.5 (eBioscience), CD44-allophycocyanin (APC) (eBioscience), CD62L-PE (eBioscience), CD71-APC (eBioscience), podoplanin-PE (eBioscience), CD31-Alexa647 (eBioscience), B220-PE (eBioscience), CD19-BB515 (BD Bioscience). Except for the lymphatic endothelial cells identification experiment, all cells were fixed in 1% PFA in 0.1 M pH 7.4 PBS. Fluorescence data were collected with a CyAn ADP High-Performance Flow Cytometer (Dako) or a Gallios (Beckman Coulter) then analysed using FlowJo software (Treestar). To obtain accurate cells counts, single cells were gated using the height, area and the pulse width of the forward and side scatter, then cells were selected for being live cells using the LIVE/DEAD Fixable Dead Cell Stain Kit per the manufacturer's instructions (Invitrogen). The cells were then gated for the appropriate markers for cell type (Extended Data Figs 3, 9). Experiments were performed on meninges

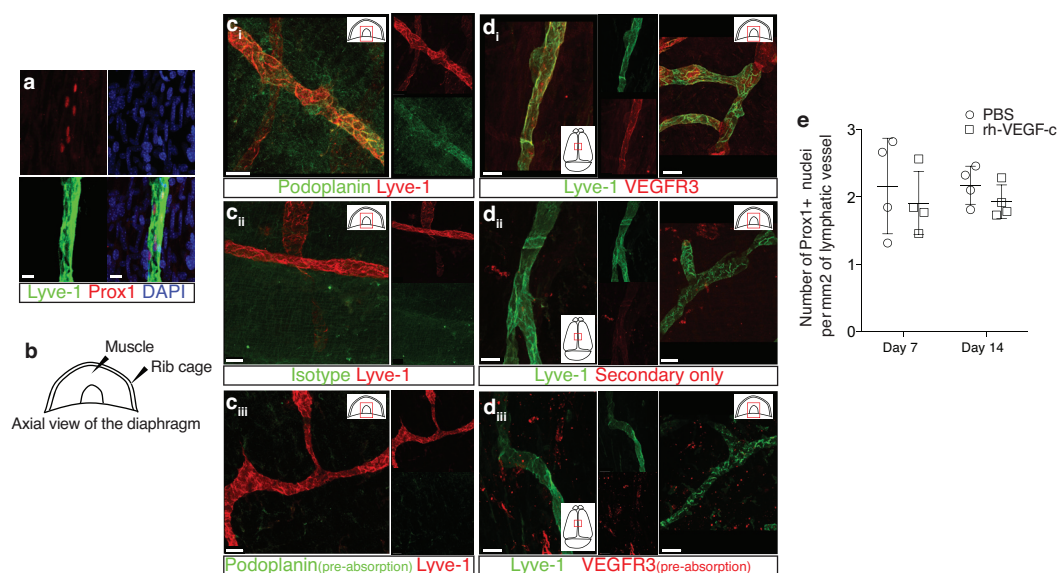
from $n = 3$ mice per group. Data processing was done with Excel and statistical analysis was performed using GraphPad Prism.

Deep cervical lymph node resection, ligation and sham surgery. Eight-week old mice were anaesthetized with ketamine/xylazine, shaved at the neck and cleaned with iodine and 70% ethanol, and an ophthalmic solution was put on the eyes to prevent drying. An incision was made midline 5 mm superior to the clavicle. The sternocleidomastoid muscle (SCM) was retracted, and the deep cervical lymph node was removed with forceps. For the ligation experiment, the collecting lymphatic vessels anterior to the deep cervical lymph nodes were ligated using a nylon suture (9-0 Ethilon black 6"VAS100-4). Sham-operated mice received the incision and had the SCM retracted, but were not ligated or the lymph nodes were not removed. Mice were then sutured and allowed to recover on a heating pad until responsive. Post-surgery, mice were given analgesic in the drinking water: 50 mg l⁻¹ for 3 days post-surgery and 0.16 mg for the next 2 weeks.



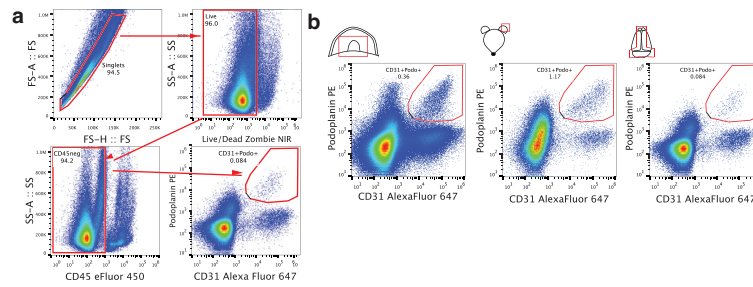
Extended Data Figure 1 | Meningeal immunity and lymphatic vessels in the dural sinuses. **a**, Representative image of CD31 staining in whole-mount meninges (scale bar, 2,000 μ m). **b**, Representative images of T cells (CD3e, arrowheads) in the dura/arachnoid, pia, dural sinuses, and choroid plexus (scale bars, 70 μ m). **c**, Quantification of T-cell density in different meningeal compartments (mean \pm s.e.m.; $n = 6$ animals each group from 2 independent experiments; *** $P < 0.001$; Kruskal–Wallis test with Dunn’s post hoc test). **d**, Quantification of MHCII-expressing cells in different meningeal compartments (mean \pm s.e.m.; $n = 6$ animals each group from 2 independent experiments; *** $P < 0.001$; Kruskal–Wallis test with Dunn’s post hoc test). **e**, Adult mice were injected i.v. with 100 μ l of DyLight 488 lectin 5 min before euthanasia to enable labelling of the cardiovascular system. Meninges were harvested and stained with anti-CD3e. Representative orthogonal images of T-cell localization in the lumen (white arrowhead) and outside of the sinus (yellow arrowhead; $n = 2$ mice; scale bar, 70 μ m). **f**, Adult mice were injected

i.v. with 10 μ g of FITC-conjugated anti-CD45 antibody or FITC-conjugated isotype antibody. Meninges were harvested one hour after the injection and labelled with anti-CD3e. Representative images of CD3e immunolabelling around dural sinuses are shown. CD45-positive cells do not co-localize with CD3e⁺ cells (**a**), suggesting an abluminal localization of the latter ($n = 2$ mice each group; scale bars, 20 μ m). **g**, Representative 3D reconstruction of the lymphatic vessels localization around the superior sagittal sinus. Adult mice were injected i.v. with 100 μ l of DyLight 488 lectin 5 min before euthanasia in order to stain the cardiovascular system. Meninges were harvested and labelled with anti-Lyve-1. The lack of lectin staining in the Lyve-1-positive meningeal lymphatic vessels suggests that they are independent of the cardiovascular system ($n = 3$ mice; scale bars, left, 50 μ m and right, 120 μ m). The mounting of the whole meninges results in the flattening of the sinus, thus it does not appear tubular.



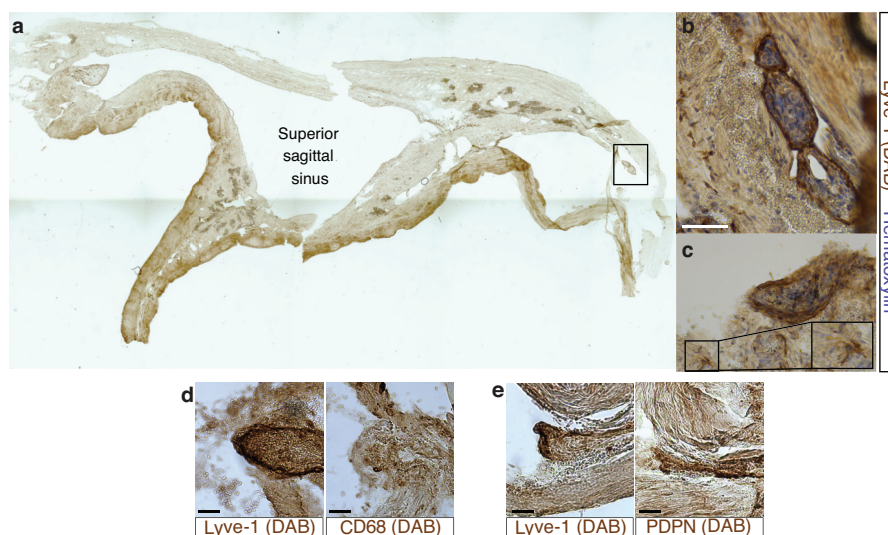
Extended Data Figure 2 | Identification, characterization and validation of the expression of classical lymphatic endothelial cell markers by the meningeal lymphatic vessels. **a**, Representative images of Prox1 labelling on meningeal Lyve-1 expressing vessels ($n = 3$ mice; scale bars, 10 μm). **b**, Schematic representation of the whole-mount dissection of the diaphragm. **c**, Characterization of the specificity of the podoplanin antibody. Representative images of whole-mount diaphragm labelled with anti-Lyve-1 and anti-podoplanin (**c_i**), control isotype (**c_{ii}**) or the anti-podoplanin pre-

incubated overnight with a saturated concentration of recombinant podoplanin protein (**c_{iii}**; scale bars, 20 μm). **d**, Characterization of the specificity of the VEGFR3 antibody. Representative images of whole-mount diaphragm and dura mater labelled with anti-Lyve-1 and anti-VEGFR3 (**d_i**), secondary antibody only (**d_{ii}**), or the anti-VEGFR3 pre-incubated overnight with a saturated concentration of recombinant VEGFR3 protein (**d_{iii}**; scale bars, 20 μm). **e**, Quantification of the number of Prox1⁺ nuclei per mm² of lymphatic vessel (mean \pm s.e.m.; $n = 4$ animals each group).



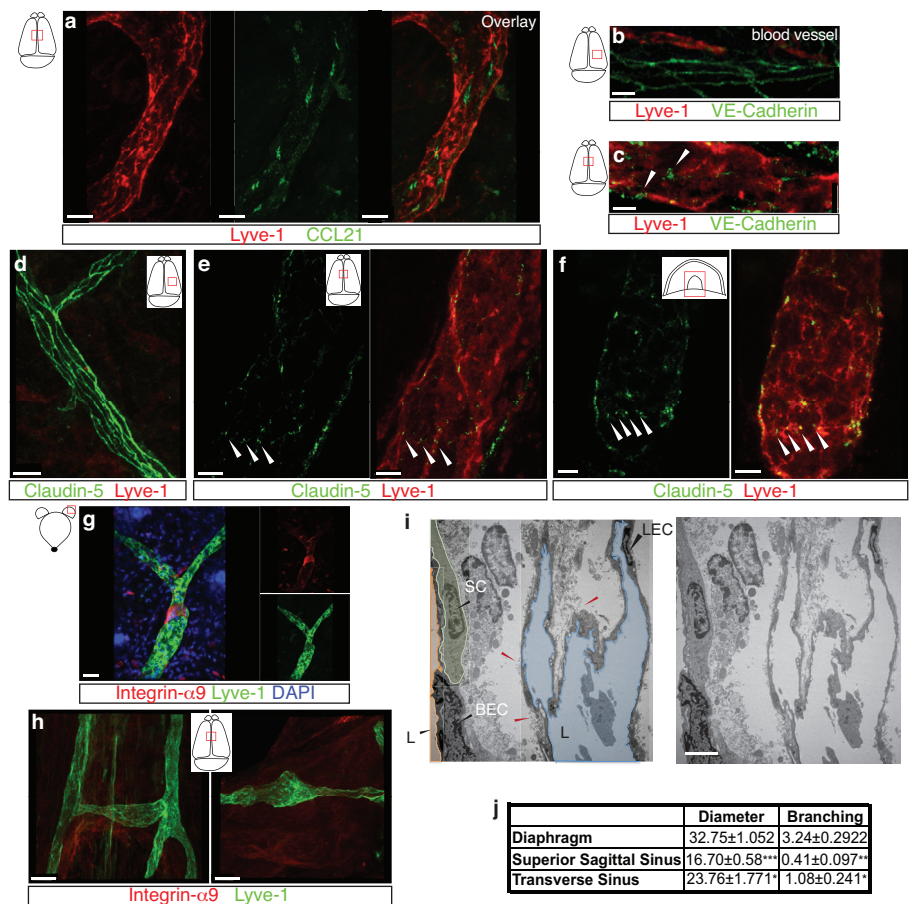
Extended Data Figure 3 | Identification of the meningeal lymphatic endothelial cell population by flow cytometry. FACS analysis of the lymphatic endothelial cells in diaphragm, skin (ear), and dural sinuses. **a**, Gating strategy employed to identify lymphatic endothelial cells (CD31⁺podoplanin⁺).

Lymphatic endothelial cells are identified as singlet, live cells, CD45⁻ and CD31⁺podoplanin⁺. **b**, Representative dot plots for lymphatic endothelial cells (CD31⁺podoplanin⁺) in the diaphragm, skin, and dura mater of adult mice.



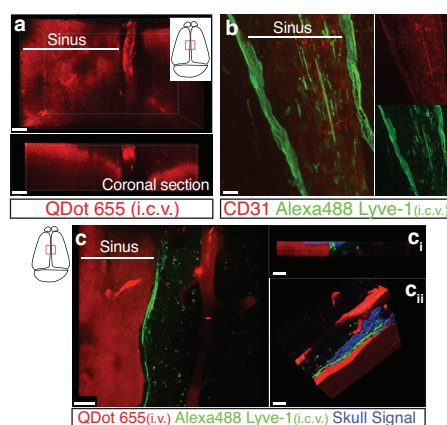
Extended Data Figure 4 | Pilot identification of lymphatic vessels in human dura. **a**, Representative image of a formalin-fixed coronal section of human superior sagittal sinus. **b**, **c**, Representative images of Lyve-1 staining on coronal section of human superior sagittal sinus (scale bar, 100 μ m). The box in **c** highlights the presence of Lyve-1-expressing macrophages in human

meninges, as seen in mice. **d**, Representative images of Lyve-1 and CD68 staining of coronal sections of human superior sagittal sinus. Note the absence of CD68 positivity on Lyve-1 positive structures (scale bars, 50 μ m). **e**, Representative images of podoplanin and Lyve-1 staining of coronal sections of human superior sagittal sinus (scale bars, 50 μ m).



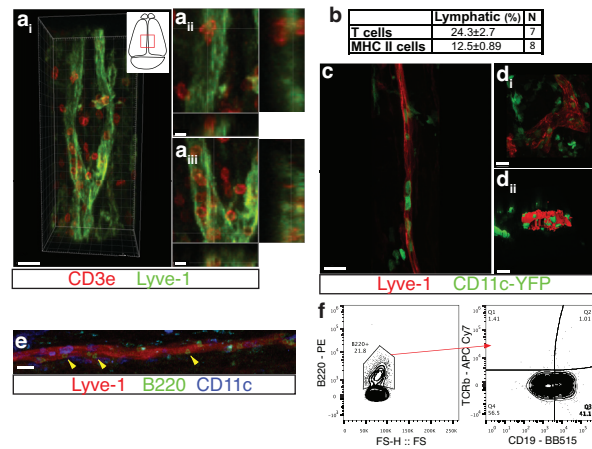
Extended Data Figure 5 | Initial lymphatic features of meningeal lymphatic vessels. **a**, Representative images of CCL21 and Lyve-1 labelling of the meningeal lymphatic vessels (scale bars, 10 μ m). **b**, **c**, Representative images of VE-Cadherin and Lyve-1 staining on meningeal blood vessels (**b**) and meningeal lymphatic vessels (**c**); arrowheads point to the VE-Cadherin aggregates; scale bars, 10 μ m). **d**–**f**, Representative images of Claudin-5 and Lyve-1 staining on meningeal blood (**d**) and lymphatic (**e**) vessels, and diaphragm lymphatic vessels (**f**); arrowheads point to Claudin-5 aggregates (scale bars, 10 μ m). **g**, **h**, Representative images of integrin- α 9 and Lyve-1 labelling on skin (**g**; ear) and meninges whole mount (**h**). Scale bars, 40 μ m. No integrin- α 9 expressing valves were detected in the meningeal lymphatic vessels.

i, Representative low power micrographs (transmission electron microscopy) of the meningeal lymphatic vessels (scale bar, 2 μ m); L, lumen; SC, supporting cell; LEC, lymphatic endothelial cell; BEC, sinusal endothelial cell. Red arrowheads point to anchoring filaments. **j**, Table summarizing morphological features of the lymphatic network in different regions of the meninges and the diaphragm. Diameters are expressed in μ m and branching as number of branches per mm of vessel; (mean \pm s.e.m.; $n = 4$ animals each group from 2 independent experiments, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$; two-way ANOVA with Bonferroni post hoc test). For statistics, the presented comparisons were between the diaphragm and the superior sagittal sinus and between the superior sagittal sinus and the transverse sinuses.



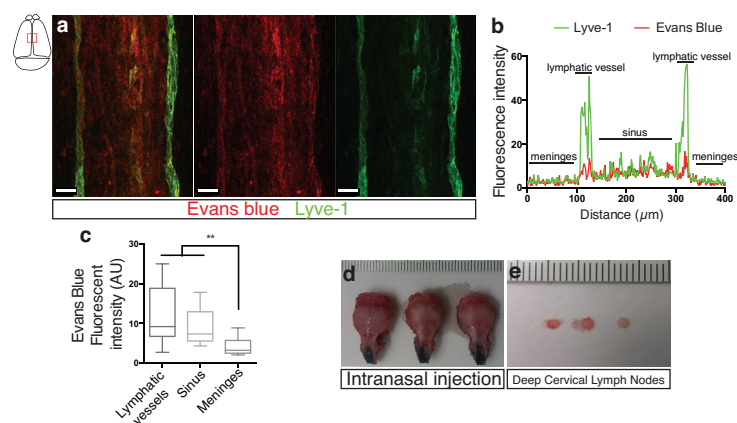
Extended Data Figure 6 | Drainage of cerebrospinal fluid into the meningeal lymphatic vessels. **a**, Representative z-stack of QDot655 filled cerebrospinal fluid drainage both in the blood vasculature (sinus) and in the meningeal lymphatic vessels after i.c.v. injection (scale bar, 20 μ m). **b**, Representative images of CD31 and Lyve-1 immunostaining on whole-mount meninges. Adult mice were injected i.c.v. with 2.5 μ g of Alexa488 conjugated anti-Lyve-1 antibody. Thirty minutes after the injection, the meninges were harvested and stained with anti-CD31. Injected *in vivo*, the

Lyve-1 antibody illuminates the lymphatic vessels (scale bar, 20 μ m). **c**, Representative z-stack of the superior sagittal sinus of adult mice injected i.v. with QDot655 and i.c.v. with Alexa488 conjugated anti-Lyve-1 antibody. **c_i**, Coronal section of the z-stack presented in panel **c**. The signal from the remaining skull and/or collagen-rich structure above the meninges was recorded (blue). **c_{ii}**, 3D reconstruction of the z-stack presented in panel **c** showing the localization of the meningeal lymphatic vessels under the skull (scale bars, 50 μ m).



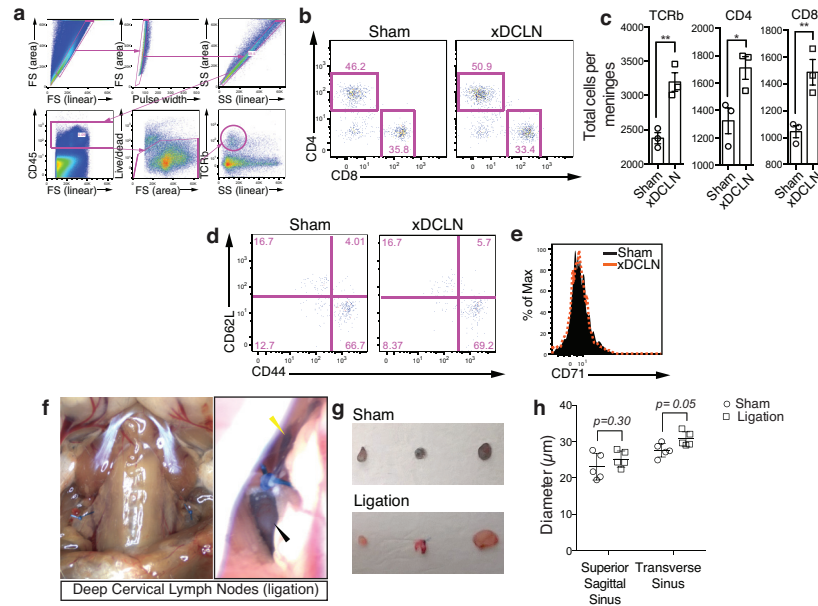
Extended Data Figure 7 | Meningeal lymphatic vessels carrying immune cells. **a**, Representative images of T cells (CD3e) and lymphatic endothelial cells (Lyve-1) on dural sinuses (scale bar, 20 μ m). **a_{ii}**–**a_{iii}**, Orthogonal sections representing T-cell localization around **a_{ii}** and within **a_{iii}** the Lyve-1 structures (scale bars, 5 μ m). **b**, Quantification of the sinusal T cells and MHCII-expressing cells within the lymphatic vessels (mean \pm s.e.m., $n = 7$ –8 mice from 3 independent experiments). **c**, **d**, Representative images of Lyve-1

staining on dural meninges from CD11c^{YFP} mice (scale bars, 20 μ m). CD11c-positive cells (most probably dendritic cells) can be found inside the meningeal lymphatic vessels. **e**, Representative image of B220⁺ cells and lymphatic endothelial cells (Lyve-1) immunolabelling in the meninges (yellow arrowheads indicate B220⁺CD11c[−] cells; scale bar, 20 μ m). **f**, Representative dot plots of B220⁺ cells (gated on singlets, live, CD45⁺) within the dural sinuses expressing CD19; ~40% of the B220⁺ cells express CD19.



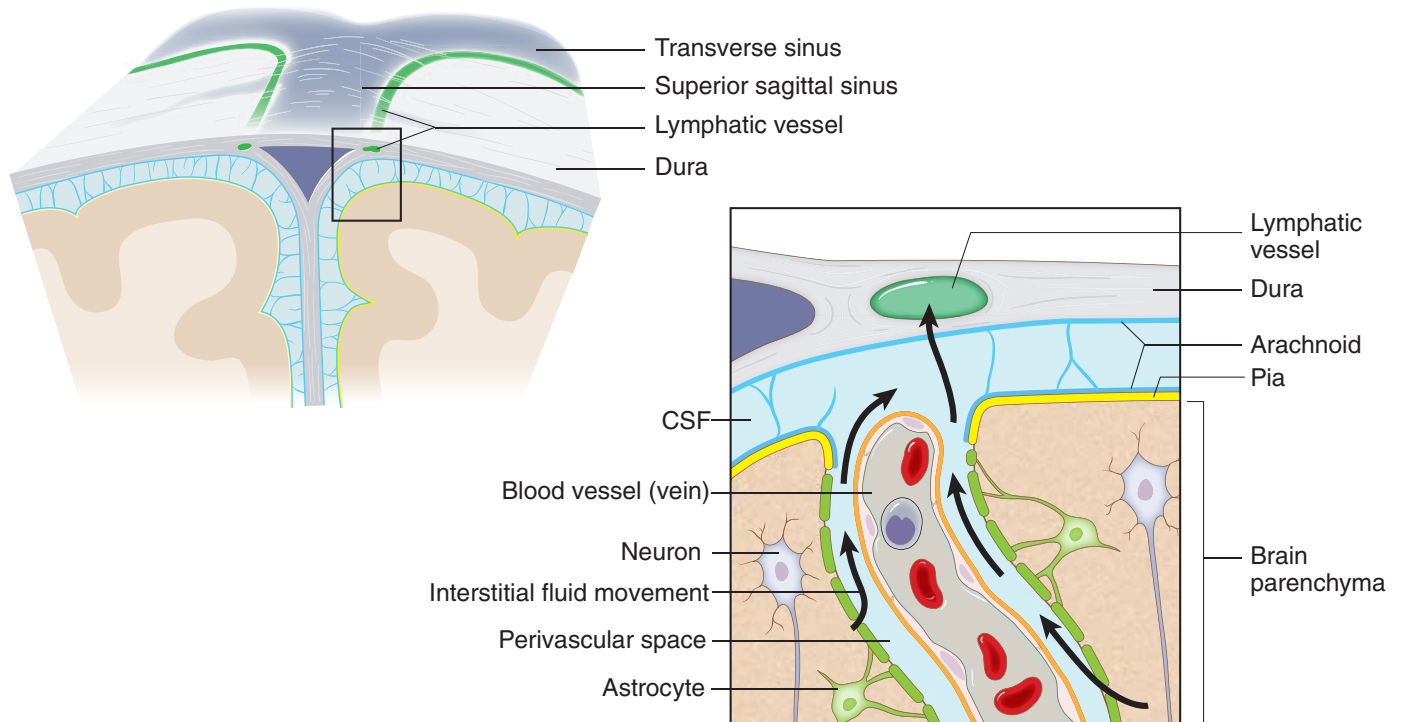
Extended Data Figure 8 | Drainage of Evans blue from the meningeal lymphatic vessels but not the nasal mucosa into the deep cervical lymph nodes. **a–c**, Adult mice were injected i.c.v. with 5 μl of 10% Evans blue. The meninges were harvested 30 min after injection and Evans blue localization was assessed by confocal microscopy. **a**, Representative images of Evans blue localization in both the sinus and the meningeal lymphatic vessels ($n = 9$ mice; scale bars, 40 μm). **b**, Representative profile of Evans blue and Lyve-1 relative fluorescence intensity on a cross-section of the image presented in **a**.

c, Quantification of the average intensity of Evans blue in the sinus, the lymphatic vessels and the meninges of adult mice (mean \pm s.e.m., $n = 16$ analysed fields from 4 independent animals; $**P < 0.01$, Kruskal–Wallis with Dunn’s multiple comparisons test). **d**, **e**, Adult mice were injected intranasally with 5 μl of 10% Evans blue. The successful targeting of the nasal mucosa (**d**) and the lack of accumulation of Evans blue in the deep cervical lymph nodes (**e**) 30 min after the injection are demonstrated.



Extended Data Figure 9 | Effects of deep cervical lymph node resection and of the lymphatic vessels ligation on the meningeal immune compartment. **a–e**, The deep cervical lymph nodes were resected (xDCLN) or sham-operated. Three weeks after resection, the meninges were harvested, single cells isolated, and analysed for T-cell content by flow cytometry. **a**, Gating strategy to analyse meningeal T cells. Meningeal T cells are selected for singlets, CD45⁺, live cells and TCRβ⁺. **b**, Representative dot plot for CD8⁺ and CD4⁺ T cells in meninges of sham and xDCLN mice. **c**, Quantification of total T cells (TCRβ⁺), CD4⁺ and CD8⁺ in the meninges of xDCLN and sham mice (mean ± s.e.m.; $n = 3$ animals each group; * $P = 0.018$; ** $P = 0.006$ (CD8), $P = 0.003$ (TCRb); Student's t -test; a representative experiment, out of two independently performed, is presented). **d**, Representative expression of CD62L and CD44 by CD4⁺ T cells phenotype in sham and xDCLN mice ($n = 3$ mice per group). **e**, Representative histogram for CD71 expression by meningeal CD4⁺ T cells in

sham and xDCLN mice ($n = 3$ mice per group). **f**, Representative images of the ligation surgery. To highlight the lymph vessels, Evans blue was injected i.c.v. before the surgery. Black arrowhead points to the node, yellow arrowhead points to the ligated Evans blue-filled vessels. **g**, Sham-operated or ligated animals were injected i.c.v. with 5 μl of 10% Evans blue. The deep cervical lymph nodes were harvested 30 min after the injection and analysed for Evans blue content. Representative images of the Evans blue accumulation in the deep cervical lymph nodes of the sham-operated and ligated animals are presented. **h**, Quantification of the meningeal lymphatic vessel diameter in the superior sagittal sinus and the transverse sinuses in sham mice and after ligation of the collecting lymphatic vessels (mean ± s.e.m., $n = 5$ mice per group from 2 independent experiments; two-way ANOVA with Bonferroni post hoc test).



Extended Data Figure 10 | Connection between the glymphatic system and the meningeal lymphatic system. A schematic representation of a connection between the glymphatic system, responsible for collecting of the interstitial

fluids from within the central nervous system parenchyma to cerebrospinal fluid, and our newly identified meningeal lymphatic vessels.

HDL-bound sphingosine-1-phosphate restrains lymphopoiesis and neuroinflammation

Victoria A. Blaho^{1,2}, Sylvain Galvani¹, Eric Engelbrecht¹, Catherine Liu¹, Steven L. Swendeman¹, Mari Kono³, Richard L. Proia³, Lawrence Steinman⁴, May H. Han⁴ & Timothy Hla^{1,2}

Lipid mediators influence immunity in myriad ways. For example, circulating sphingosine-1-phosphate (S1P) is a key regulator of lymphocyte egress^{1,2}. Although the majority of plasma S1P is bound to apolipoprotein M (ApoM) in the high-density lipoprotein (HDL) particle³, the immunological functions of the ApoM-S1P complex are unknown. Here we show that ApoM-S1P is dispensable for lymphocyte trafficking yet restrains lymphopoiesis by activating the S1P₁ receptor on bone marrow lymphocyte progenitors. Mice that lacked ApoM (*ApoM*^{-/-}) had increased proliferation of Lin⁻ Sca-1⁺ cKit⁺ haematopoietic progenitor cells (LSKs) and common lymphoid progenitors (CLPs) in bone marrow. Pharmacological activation or genetic overexpression of S1P₁ suppressed LSK and CLP cell proliferation *in vivo*. ApoM was stably associated with bone marrow CLPs, which showed active S1P₁ signalling *in vivo*⁴. Moreover, ApoM-bound S1P, but not albumin-bound S1P, inhibited lymphopoiesis *in vitro*. Upon immune stimulation, *ApoM*^{-/-} mice developed more severe experimental autoimmune encephalomyelitis⁵, characterized by increased lymphocytes in the central nervous system and breakdown of the blood-brain barrier. Thus, the ApoM-S1P-S1P₁ signalling axis restrains the lymphocyte compartment and, subsequently, adaptive immune responses. Unique biological functions imparted by specific S1P chaperones could be exploited for novel therapeutic opportunities.

S1P, a bioactive lysophospholipid mediator, interacts with vertebrate-specific S1P receptors to regulate various physiological functions⁶. Most cells express one or more of the G-protein-coupled S1P receptors (S1P₁₋₅), which mediate cellular responses such as cell migration, adhesion and survival⁷. A unique feature of this signalling system is the enrichment of the ligand, S1P, in lymph and blood compared to interstitial fluids^{1,2}. The majority (~65%) of plasma S1P is complexed with ApoM, whereas the remainder is found in the lipoprotein-free fraction, presumably associated with albumin³. Circulating ApoM is predominantly associated with a specific population of HDL particles (ApoM⁺ HDL)^{8,9}. S1P bound to ApoM⁺ HDL maintains pulmonary vascular barrier function and migration of endothelial cells *in vitro*^{3,10}.

Activation of the S1P₁ receptor on immune cells residing in secondary lymphoid organs and the thymus is critical for lymphocyte egress into circulation^{12,11}. FTY720 (fingolimod, trade name Gilenya), the first oral treatment for relapsing-remitting multiple sclerosis, targets this pathway, inducing lymphopenia and diminishing infiltration of autoreactive lymphocytes into the central nervous system (CNS)¹². Emerging evidence indicates that S1P₁ also regulates T-cell differentiation, including T helper 17 and T helper 1/regulatory T cell balance^{13,14}. However, the role of S1P signalling in lymphocyte development, homeostasis and activation remains unclear. Additionally, given the poor aqueous solubility of S1P, it is unknown whether chaperones impart specificity to S1P signals in lymphocytes, resulting in distinct immune responses.

We determined whether ApoM levels affected the S1P concentrations in blood and lymph, which are known to be important for lymphocyte egress. Although plasma S1P was decreased by ~60% (ref. 3), lymph S1P concentrations were not changed in *ApoM*^{-/-} mice (Extended Data Fig. 1a). ApoM in lymph was estimated to be approximately half of plasma levels (Extended Data Fig. 1b). Albumin

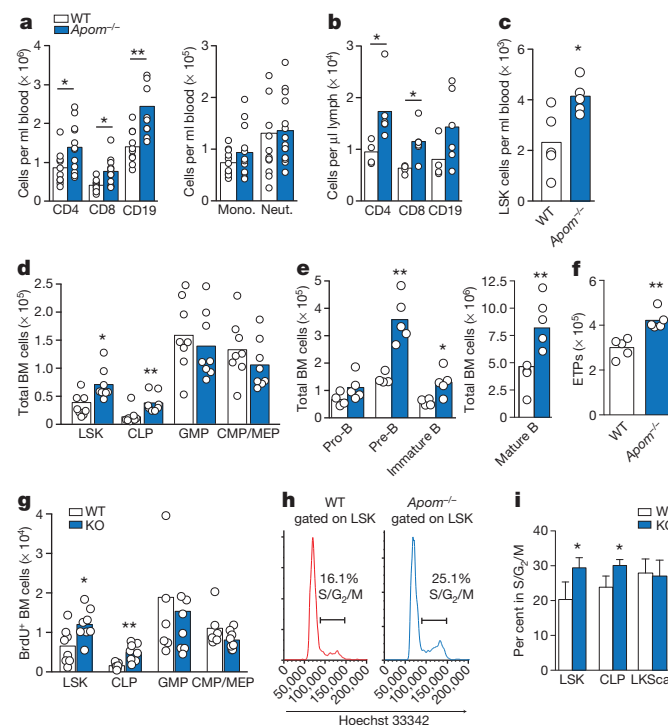


Figure 1 | Increased lymphocytes and their progenitors in *ApoM*^{-/-} mice.

a, b, CD4⁺ and CD8⁺ T cells, CD19⁺ B cells, monocytes (Mono.) and neutrophils (Neut.) in blood (**a**) or lymph (**b**) from wild-type (WT; white) and *ApoM*^{-/-} (blue) mice. **c**, LSK cells in blood of wild-type and *ApoM*^{-/-} mice. **d**, Stem and progenitor cell populations in bone marrow (BM) of wild-type and *ApoM*^{-/-} mice. **e**, Pro-, pre-, immature, and mature B-cell populations in bone marrow of wild-type and *ApoM*^{-/-} mice. **f**, ETPs in thymuses of wild-type and *ApoM*^{-/-} mice. **g**, Proliferation of progenitors in bone marrow of wild-type and *ApoM*^{-/-} mice determined by BrdU incorporation. KO, knockout. **a–g**, Circles indicate values from individual mice and bars represent means. Data are compiled from two (**b, c, e, f**) or four (**a, d, g**) independent experiments. **h**, Representative histograms of Hoechst-33342-stained LSK bone marrow cells from wild-type and *ApoM*^{-/-} mice demonstrating entry into cell cycle. Numbers above bars represent percentage of cells in S/G₂/M phase. **i**, Quantification of cycling LSK, CLP and LKSc⁺ cell subsets in wild-type and *ApoM*^{-/-} mice. Bars represent means ± standard deviation (s.d.). *n* = 4. Graphs are representative of at least two experiments. **P* < 0.05, ***P* < 0.005 as compared to wild type.

¹Center for Vascular Biology, Department of Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, New York 10065, USA. ²Brain and Mind Research Institute, Weill Medical College of Cornell University, New York, New York 10065, USA. ³Genetics of Development and Disease Branch, National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Bethesda, Maryland 20892, USA. ⁴Department of Neurology and Neurological Sciences, Stanford University, Stanford, California 94305, USA.

concentrations in blood and lymph were similar in wild-type and *Apom*^{-/-} mice (Extended Data Fig. 1c). Analysis of peripheral blood revealed a surprising increase of CD4⁺ and CD8⁺ T cells and CD19⁺ B cells in *Apom*^{-/-} mice (Fig. 1a), whereas circulating monocyte and neutrophil numbers were similar. Numbers of CD4⁺, CD8⁺ and CD19⁺ cells were also increased in lymph (Fig. 1b). Lymphocytosis was not caused by a loss of endothelial-cell S1P₁ signalling, since inducible endothelial-cell-specific deletion of *S1pr1* (S1P₁ ECKO)¹⁵ did not affect blood lymphocyte numbers (Extended Data Fig. 1d). In contrast, global knockout of *S1pr1* resulted in severe lymphopenia, consistent with a requirement for S1P₁ in lymphocyte egress from secondary lymphoid organs and the thymus (Extended Data Fig. 1d)¹. While examination of lymph nodes (brachial and inguinal) revealed similar lymphocyte numbers in *Apom*^{-/-} mice compared to wild-type (Extended Data Fig. 2a), thymuses of *Apom*^{-/-} mice contained significantly more CD4⁺CD8⁺ double-positive and CD4⁺ or CD8⁺ single-positive cells (Extended Data Fig. 2b). B-cell populations in spleens of *Apom*^{-/-} mice were slightly increased but there were no differences in the T-cell populations or spleen weights (Extended Data Fig. 2c, d). Surface expression of lymphocyte activation markers CD69 and CD62L were unchanged in the lymph nodes, thymus and spleen (Extended Data Fig. 3a–c). Administration of anti-integrin antibodies, which block lymphocyte entry into lymph nodes, had similar effects on wild-type and *Apom*^{-/-} lymph node cell numbers (Extended Data Fig. 3d), implying that ApoM⁺ HDL is not critical for lymphocyte egress.

FTY720, which induces internalization of S1P₁, induces lymphopenia^{1,12}. In both wild-type and *Apom*^{-/-} mice, administration of FTY720 resulted in marked lymphopenia in blood and lymph 2 h after administration, with similar retention patterns of increased CD4⁺, CD8⁺ and CD19⁺ cells in lymph nodes and spleen (Extended Data Fig. 4a–d). Double-negative thymocytes were decreased whereas double-positive and single-positive cells increased in thymuses of both wild-type and *Apom*^{-/-} mice (Extended Data Fig. 4e). A similar degree of lymphopenia was seen using two S1P₁-selective agonists,

AUY954 and SEW2871 (Extended Data Fig. 4f, g)^{16,17}. Collectively, these data suggest that lymphocyte trafficking out of the thymus and secondary lymphoid organs into blood and lymph is not dependent on ApoM⁺ HDL.

To determine the cause of lymphocytosis seen in *Apom*^{-/-} mice, we examined haematopoietic cell populations in blood and bone marrow (Extended Data Fig. 5a–c). LSK cells, a designation encompassing several distinct haematopoietic stem and progenitor populations, were more abundant in blood and bone marrow of *Apom*^{-/-} mice than wild-type mice (Fig. 1c). The bone marrow of *Apom*^{-/-} mice also contained increased numbers of CLPs (Lin⁻ Flt3⁺ IL7Rα⁺ cKit⁺ Sca1⁺), whereas granulocyte macrophage progenitors (GMPs; LKScal-1⁻ CD34⁺ FcγRII/III^{hi}), common myeloid progenitors (CMPs; LKScal-1⁻ CD34⁺ FcγRII/III^{lo}) and megakaryocyte/erythrocyte progenitors (MEPs; LKScal-1⁻ CD34⁻ FcγRII/III^{lo}) were unchanged (Fig. 1d)^{18–20}. Pre-, immature and mature B cells in the bone marrow (Fig. 1e) and early thymic progenitors (ETPs)/double-negative 1 cells (DN1; CD4⁻ CD8⁻ CD3⁻ cKit⁺ CD44⁺ CD25⁻) in the thymus were increased in *Apom*^{-/-} mice (Fig. 1f). Although S1P may have a role in LSK recirculation from tissues to the bone marrow¹⁹, B-cell bone marrow egress, which is insensitive to pertussis toxin, is probably not involved in the lymphocytosis seen in *Apom*^{-/-} mice²¹. Two and a half hours after bromodeoxyuridine (BrdU) injection, *Apom*^{-/-} LSK cells and CLPs had significantly more BrdU incorporation compared to wild-type cells (Fig. 1g and Extended Data Fig. 5d). No measurable difference was detected in myeloid progenitors. Cell cycle analysis demonstrated increased fractions of LSK cells and CLPs, but not LKScal-1 cells in S/G2/M phase in *Apom*^{-/-} bone marrow (Fig. 1h, i)¹⁸. These data suggested that ApoM restrained cell cycle entry specifically in bone marrow progenitors destined for lymphoid lineages.

Genetic deletion of the cholesterol transporters ATP-binding cassette A1 (ABCA1) and ABCG1, which interact with HDL to export cellular cholesterol, results in significant increases in LSK and myeloid progenitor cells^{18,22}. ABCG1 single knockout²³ or deletion of scavenger

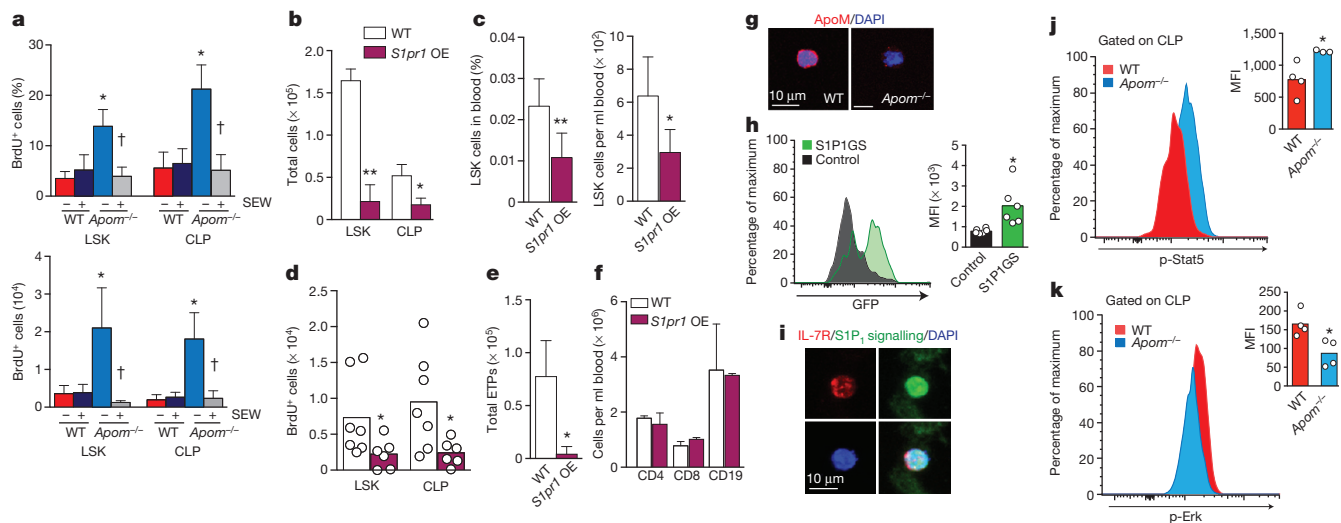


Figure 2 | ApoM-S1P-S1P₁ signalling suppresses LSK- and CLP-cell-subset expansion. **a**, Per cent BrdU incorporation (top) and total BrdU⁺ (bottom) LSK and CLP cells in wild-type (WT) and *Apom*^{-/-} bone marrow 24 h after treatment with SEW2871 (SEW). *n* = 6. **P* < 0.05 versus wild type; †*P* < 0.05 versus vehicle-treated control. **b**, **c**, LSK and CLP cells in bone marrow (**b**) and percentage and total number of LSK cells (**c**) in blood of *S1pr1* OE mice and wild-type littermates. *S1pr1* OE, *n* = 6; wild type, *n* = 7. **d**, BrdU incorporation by bone marrow LSK and CLP cells of wild-type and *S1pr1* OE mice. **e**, **f**, ETPs (**e**) or blood CD4⁺, CD8⁺ and CD19⁺ cells (**f**) from wild-type and *S1pr1* OE mice. *S1pr1* OE, *n* = 6; wild type, *n* = 7. **g**, Immunofluorescence of ApoM (red) bound to the surface of Lin⁻ Sca-1⁺ (LSK and CLP) cells from wild-type or *Apom*^{-/-} bone marrow. 4',6-Diamidino-2-phenylindole (DAPI) nuclear counterstain, blue. **h**, Representative histogram and quantitative median

fluorescence intensity (MFI) of GFP expression by CLPs from bone marrow of S1P₁ GFP signalling (S1P1GS; green) or control (black) mice. **i**, Immunofluorescence of IL-7R (red) expression and GFP (green) by CLPs. DAPI nuclear counterstain, blue. **j**, **k**, Representative flow cytometry histograms of phosphorylated (p)-Stat5 (**j**) and p-Erk1/2 (**k**) in wild-type (red) and *Apom*^{-/-} (blue) CLPs. MFI of p-Stat5 and p-Erk1/2 in CLPs of 3–4 wild-type or *Apom*^{-/-} mice are shown in inset. **d**, **h**–**k**, Bars represent means and circles represent values from individual animals. **b**, **c**, **e**, **f**, Bars represent means ± s.d. **P* < 0.05; ***P* < 0.005 versus wild type or control. **a**–**c**, **e**, **f**, **h**, **j**, **k**, Data are combined from two experiments. **d**, Data are combined from three experiments. **g**, **i**, Data are representative of at least four experiments.

receptor (SR)-B1 (ref. 24), a high-affinity HDL receptor, resulted in increased mature T and B cells, respectively, while not affecting their progenitors. Our findings suggested a bone marrow compartment-specific function of ApoM⁺ HDL in lymphocyte ontogeny. We next investigated whether interleukin (IL)-7, a key regulator of lymphopoiesis²⁰, was influenced by ApoM-S1P signalling. IL-7 concentrations in the bone marrow (Extended Data Fig. 5e) and CLP IL-7R α surface expression (Extended Data Fig. 5f) were similar in wild-type and *ApoM*^{-/-} mice. Thus, modulation of the IL-7-IL-7R α interaction is unlikely to be responsible for ApoM-mediated suppression of lymphopoiesis.

To examine the role of S1P in progenitor proliferation regulated by ApoM⁺ HDL, wild-type and *ApoM*^{-/-} mice were treated with SEW2871 and bone marrow cell BrdU incorporation was examined (Fig. 2a). SEW2871 treatment suppressed BrdU incorporation in *ApoM*^{-/-} LSK and CLP cells but not in their wild-type counterparts. Treatment with AUY954 or FTY720 also had similar effects (Extended Data Fig. 6a, b), suggesting that pharmacological activation of S1P₁ on LSK and CLP cells suppresses their proliferation. We next determined LSK and CLP cell numbers in the bone marrow of mice with an inducible *S1pr1* transgene (*S1pr1* OE; Extended Data Fig. 7a-c)¹⁵. Induction of S1P₁ expression led to decreased numbers of LSK and CLP cells (Fig. 2b). Blood LSK cell numbers were also reduced, ruling out increased bone marrow egress of haematopoietic progenitors (Fig. 2c). Both LSK and CLP populations showed decreased BrdU incorporation (Fig. 2d). S1P₁ overexpression also resulted in decreased ETPs in the thymus (Fig. 2e). Mature T- and B-cell numbers in the blood were unchanged and the secondary lymphoid organs had reduced cell numbers (Fig. 2f and Extended Data Fig. 7d-f), suggesting increased egress of mature lymphocytes from lymphoid organs upon S1P₁ overexpression. These pharmacological and genetic data further confirm that ApoM⁺ HDL signalling via bone marrow progenitor-cell S1P₁ restrains lymphopoiesis.

Real-time reverse transcription-polymerase chain reaction (RT-PCR) analysis found that *S1pr1* messenger RNA is expressed by CLPs and confirmed a previous report of its expression by LSK cells (Extended Data Fig. 8a)¹⁹. Immunofluorescence analysis of Lin⁻ Sca-1⁺ bone marrow cells from wild-type mice showed stable association of ApoM with the cell surface (Fig. 2g). To determine if S1P₁ is active in bone marrow progenitor cells *in vivo*, we used S1P₁ green fluorescent protein (GFP) signalling mice, which show a nuclear GFP signal upon receptor activation (Extended Data Fig. 8b-f)⁴, and found that CLP and LSK cells show active S1P₁ signalling (Fig. 2h, i).

We next examined the signal transducer and activator of transcription 5 (Stat5) pathway, which is activated by IL-7R signalling and regulates lymphocyte development, proliferation and differentiation^{25,26}. Phosphoflow analysis (Extended Data Fig. 8g, h) revealed significantly increased phospho-Stat5 in *ApoM*^{-/-} CLPs compared to wild-type cells (Fig. 2j). Conversely, extracellular-signal-regulated kinases 1 and 2 (Erk1/2) antagonize the proliferative signals induced by Stat5 (ref. 26). CLPs from mice lacking ApoM-S1P had significantly less phospho-Erk1/2 than wild-type CLPs (Fig. 2k). Collectively, these data suggest that ApoM-S1P directly interacts with bone marrow lymphocyte progenitors and activates the S1P₁/Erk pathway to restrain their proliferation.

To test whether ApoM⁺ HDL modulated lymphocyte development directly, we used *in vitro* lymphopoiesis assays. Using Lin⁻ cells from wild-type bone marrow, we observed that the generation of B220⁺ cells (pre-, pro- and immature B lymphocytes) was strongly suppressed by ApoM⁺ HDL but not by albumin-bound S1P (Fig. 3a, b and Extended Data Fig. 9a). HDL from *ApoM*^{-/-} mice, which lack S1P (ApoM⁻ HDL) (Extended Data Fig. 9b), did not suppress lymphopoiesis (Fig. 3c). ApoM⁺ HDL-mediated suppression of lymphopoiesis was blocked by pre-incubation of HDL with an anti-S1P immunoglobulin M (IgM) but not by an irrelevant IgM (Fig. 3d). These findings

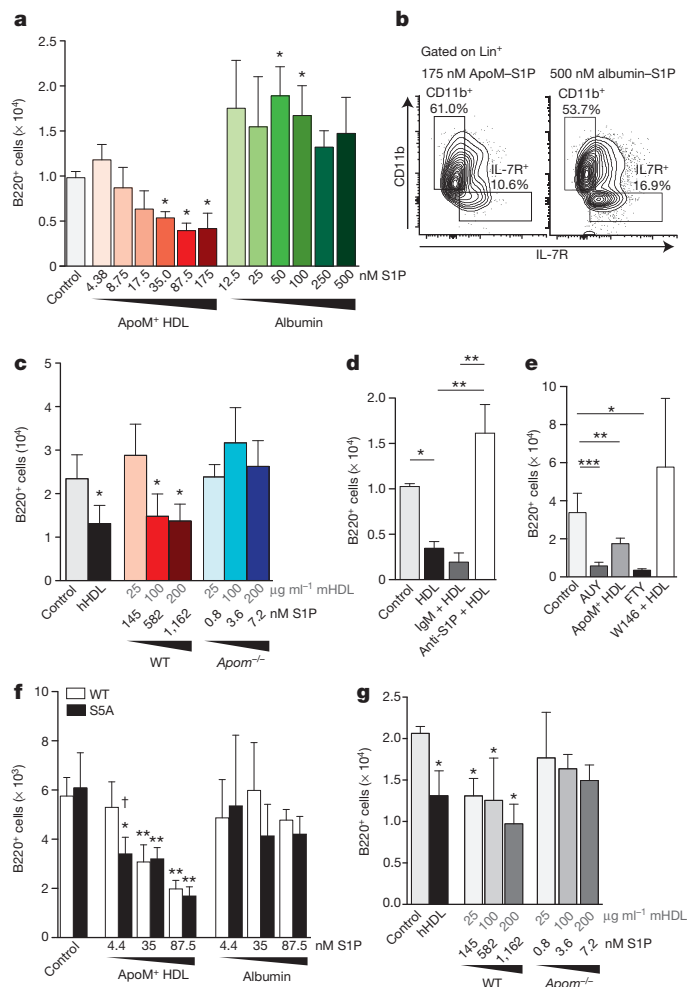


Figure 3 | ApoM-S1P activation of S1P₁ suppresses lymphocyte progenitor generation *in vitro*. **a**, B220⁺ cells generated from wild-type (WT) Lin⁻ bone marrow after incubation with increasing concentrations of ApoM⁺ HDL S1P or albumin-S1P. **b**, Representative flow cytometry plots of IL-7R α ⁺ versus CD11b⁺ cells generated in the presence of 175 nM ApoM-S1P or 500 nM albumin-S1P. **c**, B220⁺ cell numbers generated after culture in the presence of human HDL (hHDL; 100 μ g ml⁻¹), or increasing concentrations of mouse ApoM⁺ HDL (wild type) or ApoM⁻ HDL (*ApoM*^{-/-}). **d**, B220⁺ cells generated after incubation with ApoM⁺ HDL (50 μ g ml⁻¹), ApoM⁺ HDL plus IgM, or ApoM⁺ HDL plus anti-S1P. **e**, B220⁺ cells generated in the presence of vehicle, AUY954 (AUY; 10 nM), FTY720p (FTY; 10 nM), 100 μ g ml⁻¹ ApoM⁺ HDL, or W146 (100 nM) plus 100 μ g ml⁻¹ HDL. **f**, B220⁺ cells generated from wild-type or S1P₁^{S5A} (S5A) Lin⁻ bone marrow cells after incubation in the presence of increasing concentrations of ApoM⁺ HDL-S1P or albumin-S1P. **g**, Number of B220⁺ cells generated from S5A Lin⁻ bone marrow incubated with increasing concentrations of ApoM⁺ HDL (wild type) or ApoM⁻ HDL (*ApoM*^{-/-}). **a-e**, $n = 5$ (**a-c**) or 6 (**d, e**); data are compiled from two independent experiments. **f, g**, Wild type, $n = 5$; S5A, $n = 3$; data are representative of two experiments. * $P < 0.05$, ** $P < 0.005$, *** $P < 0.001$ versus control or vehicle values. † $P < 0.05$ versus wild-type value. Bars represent mean \pm s.d.

suggest that ApoM⁺ HDL-bound S1P is critical to restrain lymphocyte progenitor differentiation and proliferation.

To determine the involvement of S1P₁, cultures were incubated with AUY954 or phosphorylated FTY720 (FTY720p), resulting in significantly decreased numbers of B220⁺ cells (Fig. 3e). However, incubation of Lin⁻ bone marrow cells with W146, a selective S1P₁ antagonist, abrogated the suppressive effect of ApoM⁺ HDL. Next, we used Lin⁻ bone marrow from S1P₁^{S5A} mice, which express a mutated S1P₁ that is internalization defective, resulting in extended surface residency and sustained signalling^{13,17,27}. S1P₁^{S5A} bone marrow cells were more

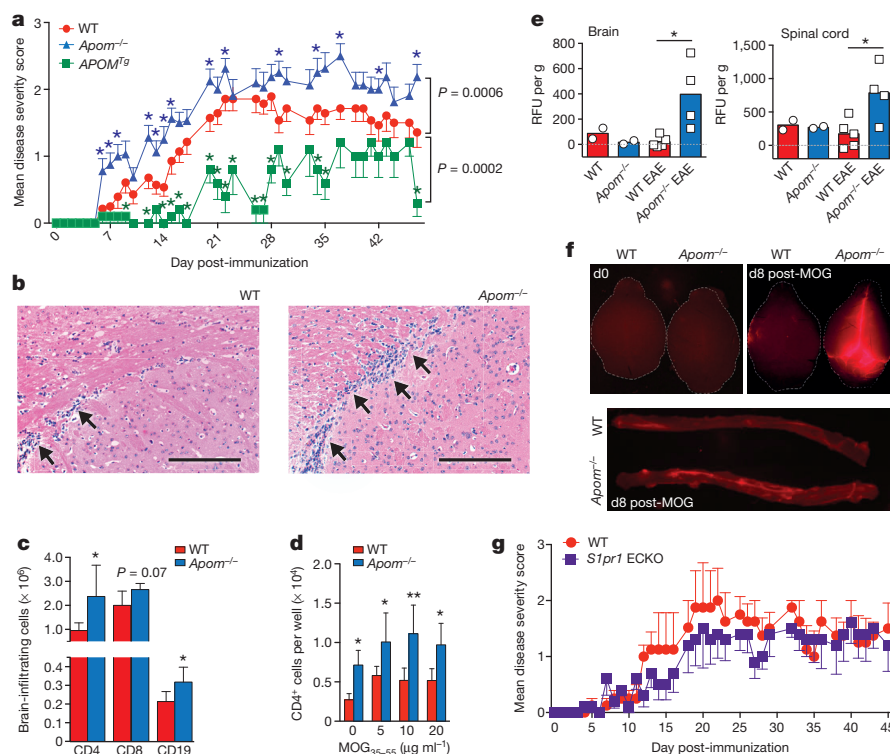


Figure 4 | ApoM expression affects EAE disease onset and severity. **a**, Mean clinical EAE scores of wild-type (WT), *Apom*^{-/-} or *APOM*^{Tg} mice. Circles represent means ± standard error of the mean (s.e.m.). Wild type, *n* = 14; *Apom*^{-/-}, *n* = 16; *APOM*^{Tg}, *n* = 5. Data from wild-type and *Apom*^{-/-} mice are representative of three independent experiments. **P* ≤ 0.05 compared to wild-type value as determined by Mann–Whitney U-test. **b**, Representative photomicrographs of haematoxylin and eosin staining of brains from wild-type or *Apom*^{-/-} mice at day 16 after MOG_{35–55} immunization. Scale bar, 100 μm. **c**, Numbers of brain-infiltrating CD4⁺, CD8⁺ and CD19⁺ cells from the CNS of wild-type and *Apom*^{-/-} EAE mice at day 24 after immunization. **d**, Number of CD4⁺ T cells from splenocytes of MOG_{35–55}-immunized wild-type or

Apom^{-/-} EAE mice after *ex vivo* restimulation with increasing concentrations of MOG_{35–55} peptide. **c**, **d**, Data are means ± s.d. **P* < 0.05, ***P* < 0.01 compared to wild type. Experiments were performed at least twice with cells from *n* = 5 per arm. **e**, Quantification of TMR-dextran fluorescence in brain and spinal cord after 16 h of circulation in wild-type or *Apom*^{-/-} mice at day 0 or day 8 after MOG_{35–55} immunization. Bars represent means and circles and squares represent values from individual mice. Data are combined from two experiments. RFU, relative fluorescence units. **f**, Representative photographs of whole brains or spinal cords quantified in **e**. **g**, Mean clinical EAE scores of *S1pr1* ECKO and wild-type littermates. Circles are means ± s.e.m. *n* = 5 per arm and data are representative of two experiments performed.

sensitive to ApoM–S1P than wild-type bone marrow, since lower S1P concentrations were sufficient to repress B220⁺ development (Fig. 3f). Despite greater S1P sensitivity, albumin–S1P was unable to suppress lymphopoiesis by S1P₁^{S5A} cells, similar to wild-type cells. ApoM⁻ HDL did not suppress B220⁺ generation from S1P₁^{S5A} bone marrow cells (Fig. 3g). Collectively, these findings support the hypothesis that ApoM⁺ HDL activation of S1P₁ on lymphocyte progenitors inhibits their proliferation.

We hypothesized that modulation of the lymphocyte compartment size by ApoM⁺ HDL could impact the magnitude of immune responses. The experimental autoimmune encephalomyelitis (EAE) model recapitulates many aspects of multiple sclerosis pathology, including increased bone marrow cell proliferative capacity, blood–brain barrier (BBB) disruption, and autoreactive immune cell CNS infiltration⁵. Interestingly, increased plasma HDL correlates with decreased acute inflammation in multiple sclerosis patients, although the mechanism and the role of specific HDL populations are unknown^{28,29}. To ascertain whether the expanded lymphocyte compartment in *Apom*^{-/-} mice could affect the extent and intensity of the adaptive immune response, we induced EAE in wild-type, *Apom*^{-/-} mice, and mice overexpressing a human *APOM* transgene (*APOM*^{Tg})³ by immunization with myelin oligoglycoprotein (MOG_{35–55}) peptide. Mice lacking ApoM developed symptoms earlier and exhibited more severe CNS pathology compared to wild-type mice (Fig. 4a). Remarkably, *APOM*^{Tg} mice developed milder symptoms, supporting a protective role for ApoM–S1P (Fig. 4a). More severe inflammation in the CNS was evident at day 16 after immunization in *Apom*^{-/-} brains

as compared to the wild-type counterparts (Fig. 4b), and increased lymphocytes were found in *Apom*^{-/-} brains at the peak of disease (Fig. 4c). Lymphocytes from MOG-immunized *Apom*^{-/-} mice were highly activated, as demonstrated by increased splenocyte proliferative responses to *ex vivo* MOG restimulation (Fig. 4d). Thus, the increase in bone marrow proliferation of lymphocyte progenitors and increased circulating lymphocytes in *Apom*^{-/-} mice resulted in an exaggerated autoimmune neuroinflammatory response, whereas overexpression of ApoM was protective.

Previously, our laboratory has reported that the loss of ApoM results in enhanced pulmonary vascular leakage³. To determine whether altered vascular barrier function was contributing to increased EAE severity, wild-type and *Apom*^{-/-} mice were injected intravenously with tetramethylrhodamine (TMR)-labelled dextran (70 kilodalton) before or 8 days after MOG_{35–55} immunization to assess BBB functionality. Before immunization, no differences in dextran extravasation were evident in the brain or spinal cord (Fig. 4e, f and Extended Data Fig. 10a). After immunization, *Apom*^{-/-} brains and spinal cords displayed significantly increased vascular leakage. This was dependent upon the immune response, since pertussis toxin alone was insufficient to induce vascular leakage (Extended Data Fig. 10b). Thus, the BBB was intact in mice lacking ApoM before MOG_{35–55} immunization, and induction of an immune response was necessary for the generation of EAE and associated BBB breakdown.

S1P₁ ECKO mice have increased pulmonary vascular leakage (Extended Data Fig. 10c); however, deletion of endothelial *S1pr1* did not influence EAE disease severity (Fig. 4g), suggesting that alterations

in vascular barrier function are not sufficient to modulate autoimmune responses in the CNS. Similarly, S1P₁ expression by astrocytes has been shown to affect EAE severity³⁰. Since neurovascular leakage would be a necessary antecedent for plasma ApoM–S1P to gain access to astrocytes, it seems unlikely that the increased disease severity seen in *ApoM*^{−/−} mice is due to direct ApoM effects on neural cells. Collectively, these data indicate that ApoM–S1P signalling in lymphocytes is a key factor in the increased EAE pathology displayed by *ApoM*^{−/−} mice.

In summary, these data reveal an unexpected mechanism by which ApoM⁺ HDL specifically restrains bone marrow lymphopoiesis by inducing S1P/S1P₁ receptor signalling in bone marrow progenitors, thereby limiting the lymphocyte compartment. These findings also suggest that unique functions of S1P are imparted by its chaperone, ApoM⁺ HDL, and that ApoM⁺ HDL-dependent S1P/S1P₁ signalling induces distinct physiological outcomes. Although dispensable for lymphocyte trafficking, ApoM–S1P attenuation of lymphopoiesis under homeostasis may be an important regulator of the lymphocyte life cycle, and therefore have a role in autoimmune responses. Our studies indicate that ApoM⁺ HDL-mediated regulation of lymphocyte ontogeny impacts the extent and intensity of autoimmune neuroinflammation in the EAE model. Furthermore, chaperone-dependent S1P signalling may allow the development of novel strategies for the control of immune responses.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 March 2014; accepted 8 April 2015.

Published online 8 June 2015.

- Cyster, J. G. & Schwab, S. R. Sphingosine-1-phosphate and lymphocyte egress from lymphoid organs. *Annu. Rev. Immunol.* **30**, 69–94 (2012).
- Schwab, S. R. *et al.* Lymphocyte sequestration through S1P lyase inhibition and disruption of S1P gradients. *Science* **309**, 1735–1739 (2005).
- Christoffersen, C. *et al.* Endothelium-protective sphingosine-1-phosphate provided by HDL-associated apolipoprotein M. *Proc. Natl Acad. Sci. USA* **108**, 9613–9618 (2011).
- Kono, M. *et al.* Sphingosine-1-phosphate receptor 1 reporter mice reveal receptor activation sites *in vivo*. *J. Clin. Invest.* **124**, 2076–2086 (2014).
- Furlan, R., Cuomo, C. & Martino, G. Animal models of multiple sclerosis. *Methods Mol. Biol.* **549**, 157–173 (2009).
- Hannun, Y. A. & Obeid, L. M. Principles of bioactive lipid signalling: lessons from sphingolipids. *Nature Rev. Mol. Cell Biol.* **9**, 139–150 (2008).
- Chun, J., Hla, T., Lynch, K. R., Spiegel, S. & Moolenaar, W. H. International Union of Basic and Clinical Pharmacology. LXXVIII. Lysophospholipid receptor nomenclature. *Pharmacol. Rev.* **62**, 579–587 (2010).
- Faber, K., Axler, O., Dahlbäck, B. & Nielsen, L. B. Characterization of ApoM in normal and genetically modified mice. *J. Lipid Res.* **45**, 1272–1278 (2004).
- Christoffersen, C. *et al.* Isolation and characterization of human apolipoprotein M-containing lipoproteins. *J. Lipid Res.* **47**, 1833–1843 (2006).
- Kimura, T. *et al.* Sphingosine 1-phosphate may be a major component of plasma lipoproteins responsible for the cytoprotective actions in human umbilical vein endothelial cells. *J. Biol. Chem.* **276**, 31780–31785 (2001).
- Allende, M. L., Dreier, J. L., Mandala, S. & Proia, R. L. Expression of the sphingosine 1-phosphate receptor, S1P₁, on T-cells controls thymic emigration. *J. Biol. Chem.* **279**, 15396–15401 (2004).
- Brinkmann, V. *et al.* Fingolimod (FTY720): discovery and development of an oral drug to treat multiple sclerosis. *Nature Rev. Drug Discov.* **9**, 883–897 (2010).
- Garris, C. S. *et al.* Defective sphingosine 1-phosphate receptor 1 (S1P₁) phosphorylation exacerbates T_H17-mediated autoimmune neuroinflammation. *Nature Immunol.* **14**, 1166–1172 (2013).
- Liu, G., Yang, K., Burns, S., Shrestha, S. & Chi, H. The S1P₁ mTOR axis directs the reciprocal differentiation of T_H1 and T_{reg} cells. *Nature Immunol.* **11**, 1047–1056 (2010).
- Jung, B. *et al.* Flow-regulated endothelial S1P receptor-1 signaling sustains vascular development. *Dev. Cell* **23**, 600–610 (2012).
- Jo, E. *et al.* S1P₁-selective *in vivo*-active agonists from high-throughput screening: off-the-shelf chemical probes of receptor interactions, signaling, and fate. *Chem. Biol.* **12**, 703–715 (2005).
- Thangada, S. *et al.* Cell-surface residence of sphingosine 1-phosphate receptor 1 on lymphocytes determines lymphocyte egress kinetics. *J. Exp. Med.* **207**, 1475–1483 (2010).
- Yvan-Charvet, L. *et al.* ATP-binding cassette transporters and HDL suppress hematopoietic stem cell proliferation. *Science* **328**, 1689–1693 (2010).
- Massberg, S. *et al.* Immunosurveillance by hematopoietic progenitor cells trafficking through blood, lymph, and peripheral tissues. *Cell* **131**, 994–1008 (2007).
- Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
- Beck, T. C., Gomes, A. C., Cyster, J. G. & Pereira, J. P. CXCR4 and a cell-extrinsic mechanism control immature B lymphocyte egress from bone marrow. *J. Exp. Med.*, (2014).
- Westerterp, M. *et al.* Regulation of hematopoietic stem and progenitor cell mobilization by cholesterol efflux pathways. *Cell Stem Cell* **11**, 195–206 (2012).
- Armstrong, A. J., Gebre, A. K., Parks, J. S. & Hedrick, C. C. ATP-binding cassette transporter G1 negatively regulates thymocyte and peripheral lymphocyte proliferation. *J. Immunol.* **184**, 173–183 (2010).
- Feng, H. *et al.* Deficiency of scavenger receptor BI leads to impaired lymphocyte homeostasis and autoimmune disorders in mice. *Arterioscler. Thromb. Vasc. Biol.* **31**, 2543–2551 (2011).
- Mazzucchelli, R. & Durum, S. K. Interleukin-7 receptor expression: intelligent design. *Nature Rev. Immunol.* **7**, 144–154 (2007).
- Clark, M. R., Mandal, M., Ochial, K. & Singh, H. Orchestrating B cell lymphopoiesis through interplay of IL-7 receptor and pre-B cell receptor signalling. *Nature Rev. Immunol.* **14**, 69–80 (2014).
- Oo, M. L. *et al.* Engagement of S1P₁-degradative mechanisms leads to vascular leak in mice. *J. Clin. Invest.* **121**, 2290–2300 (2011).
- Weinstock-Guttman, B. *et al.* Serum lipid profiles are associated with disability and MRI outcomes in multiple sclerosis. *J. Neuroinflammation* **8**, 127 (2011).
- Weinstock-Guttman, B., Zivadinov, R. & Ramanathan, M. Inter-dependence of vitamin D levels with serum lipid profiles in multiple sclerosis. *J. Neurol. Sci.* **311**, 86–91 (2011).
- Choi, J. W. *et al.* FTY720 (fingolimod) efficacy in an animal model of multiple sclerosis requires astrocyte sphingosine 1-phosphate receptor 1 (S1P₁) modulation. *Proc. Natl Acad. Sci. USA* **108**, 751–756 (2011).

Acknowledgements The authors thank C. Nathan, G. Koretzky and R. Nachman for critical comments, C. Christoffersen and L. B. Nielsen for the *ApoM*^{−/−} and *APOM*^{fl} mice and discussions, V. Brinkmann and Novartis Pharma AG for gifts of FTY720 and AUY954, M. Sanson for help with HDL purification, Y. Huang for help with the EAE model and J. McCormick for assistance with fluorescence-activated cell sorting. V.A.B. is a Leon Levy Research Fellow of the Weill Cornell Medical College Brain and Mind Institute. Research was supported by grants to V.A.B. (National Institutes of Health (NIH) F32 CA14211 and New York Stem Cell Foundation C026878), T.H. (NIH HL67330, HL70694 and HL89934, Fondation Leducq), R.L.P. (Intramural program of the National Institute of Diabetes and Digestive and Kidney Diseases, NIH and Fondation Leducq), M.H. and L.S. (Fondation Leducq), the Lipidomics Shared Resource, Hollings Cancer Center, Medical University of South Carolina (P30 CA138313) and the Lipidomics Core in the SC Lipidomics and Pathobiology COBRE (P20 RR017677).

Author Contributions V.A.B. designed and performed the experiments, analysed and interpreted data, and wrote the manuscript. S.G., E.E., C.L. and S.L.S. performed experiments and analysed data. L.S. and M.H.H. contributed to EAE studies. R.L.P. and M.K. contributed to the reporter mouse studies. T.H. supervised the overall project, designed experiments, interpreted data, and wrote the manuscript. All authors read and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.H. (tih2002@med.cornell.edu).

METHODS

Animals. *Apom*^{-/-} and *APOM*^{Tg} mice were crossed at least nine generations to the C57BL/6 genetic background^{3,31}. C57BL/6 mice purchased from Jackson Laboratory were used as wild-type controls. To generate mice that overexpress SIP₁, mice expressing the *S1pr1*^{flox/flox} transgene were crossed to mice expressing the tamoxifen-inducible *Rosa26-Cre-ER*^{T2} to yield *S1pr1*^{flox/flox} *Rosa26-Cre-ER*^{T2} animals¹⁵. To generate mice with endothelial-cell-specific deletion of *S1pr1*, mice with floxed *S1pr1* genes (*S1pr1*^{fl/fl}) were crossed with mice expressing a tamoxifen-inducible Cre under the control of the VE-cadherin gene (*Cdh5-Cre-ER*^{T2})^{15,32}. For both *S1pr1*-overexpressing mice (*S1pr1* OE) and endothelial-cell-specific deletion (ECKO), Cre-negative littermates were used as controls. To induce deletion of the floxed *S1pr1* or floxed 'stop' sequence, mice were treated with tamoxifen by oral gavage, 4 mg every third or fifth day for a total of three treatments. *S1pr1* OE mice were analysed 3–4 days after the final tamoxifen administration; ECKO mice were analysed ≥10 days after the final tamoxifen administration. SIP₁ GFP signalling mice were generated as previously described⁴. Briefly, mice expressing a transcriptional unit consisting of the SIP₁ carboxy terminus fused to tetracycline transcriptional activator (tTA) and β-arrestin fused to tobacco etch virus (TEV) protease knocked in to the endogenous SIP₁ locus were crossed with mice containing a histone–GFP reporter gene under the control of a tTA-responsive promoter. Mice expressing one allele of both the *S1pr1*Ki and the GFP alleles were considered SIP₁ GFP signalling mice. Littermates expressing only the GFP allele without the SIP₁ knock-in were considered controls. For all bone marrow studies, animals were killed and tissues collected between 12:00 and 15:00 to limit circadian effects. No randomization method was used and investigators were blinded to the genotype of animals during data analysis. The minimum number of animals per group was four and for all inducible Cre strains wild-type littermate controls were used. Male and female mice were used as age-matched controls for all studies except for EAE experiments, for which only female animals were used. No animals were excluded from the analysis. Animals were housed in a specific pathogen-free facility and provided food and water *ad libitum*. All animal protocols were approved by the Institutional Animal Care and Use Committee of Weill Cornell Medical College.

Verification of *S1pr1* mRNA overexpression and floxed 'stop' excision. After completion of the tamoxifen treatment regimen, RNA was extracted from whole bone marrow with Trizol according to the manufacturer's instructions, and TaqMan real-time multiplex RT–PCR was performed as described later to assess overexpression of *S1pr1* mRNA. For extraction of genomic DNA from the same samples, back extraction buffer (1 M Tris, pH 8.0, 4 M guanidine thiocyanate, 50 mM sodium citrate) was added to the remaining organic phase and interphase, vortexed, and incubated for 10 min at room temperature. Samples were then centrifuged at 4 °C for 15 min at 12,000g and the aqueous phase transferred to a new tube. DNA was precipitated with 400 μl isopropanol and samples centrifuged again at 4 °C for 15 min at 12,000g. Supernatant was removed, samples washed with 75% ethanol and the DNA pellet was resuspended in H₂O. PCR was performed using 20 ng of DNA, with the following primers. 5': GTTCGGCTCTGGCGTGTGA; 3': CCAGTCTCTCC TCGGAGATC; and the reaction conditions: 95 °C/30 s, 56 °C/30 s, 72 °C/1 min for 30 cycles. Amplification of the intact floxed *stop* cassette yielded a 1.4 bp fragment and, when deleted, a 308 bp fragment.

Cell and plasma collection from blood and lymph. Mice were euthanized with CO₂. Blood was recovered via terminal cardiac puncture and collected in tubes containing 35 μl 0.1 M EDTA. Five-hundred millilitres of whole blood was then removed to a new tube, centrifuged at 2,000g for 15 min, and plasma removed and stored at –20 °C. Cells were then subjected to ammonium chloride (ACK) erythrocyte lysis, washed with PBS, and stained with eFluor718 fixable viability dye (eBioscience) according to manufacturer's instructions. Lymph was recovered from the cisterna chyli, as previously described^{23,33}, and collected in 5 μl acid citrate-dextrose. Total lymph volume was recorded before the entire sample was centrifuged at 2,000g for 15 min to isolate plasma. Plasma was then removed and cells were washed with PBS before staining with the eFluor718 viability dye. After a PBS wash, blood and lymph cells were then briefly fixed for 30 min in 0.1% paraformaldehyde (PFA), washed, and stained with antibody cocktails for flow cytometry, as described later³⁴.

Tissue cell isolation. Single-cell suspensions were created from spleen, thymus and lymph nodes (two inguinal and brachial lymph nodes per mouse) by gently disaggregating tissue between frosted glass slides. Bone marrow cells were collected after removing the epiphyses and a brief centrifugation in a 1.5 ml tube. Cells were then washed with 1% FBS in PBS and subjected to ACK lysis buffer. After another PBS wash, cells were stained with eFluor718 viability dye. For some studies, such as cell sorting, cells were stained for and analysed by flow cytometry unfixed. Except for studies requiring live sorted cells, samples were gently fixed with 0.1% PFA for 10 min and then washed with 1% FBS in PBS, unless noted otherwise. Although absolute numbers were lower after fixation, this did not have

a dramatic effect on staining patterns and gave results similar to those obtained using unfixed cells.

Flow cytometry. After fixation, cells were washed with 1% FBS in PBS (flow buffer (FB)) and Fc receptors were blocked with TruStain FcX (Biolegend) in 3% FBS overnight at 4 °C. Cells were then aliquoted and stained in their respective antibody cocktails for 30 min on ice. After washing with FB, liquid counting beads (BD) were added to each sample and data were obtained on a BD LSRII (Becton Dickinson) and analysed using FlowJo software (Tree Star, Inc.).

Antibodies were purchased from BioLegend (BL), eBioscience (eB), Southern Biotech (SB), or Becton Dickinson (BD). For the determination of blood and lymph cell populations the following clones and antibody–fluorophore conjugates were used: CD4–PE (RM4-4, BL), CD8–PerCP/eFluor710 (H35-17.2; eB), CD19–eFluor450 (ID3; eB), CD11b–APC–eFluor780 (M1/70; eB), Ly6C–FITC (HK1.4; SB), Ly6G–PE–Cy7 (1A8; BL) and CD115–APC (AFS98; BL). For the determination of lymph node and splenocyte populations: CD4–APC–Cy7 (GK1.5; BL), CD8–eFluor450 (H35-17.2; eB), B220–AF700 (RA3-6B2; eB), CD62L–PerCP–Cy5.5 (MEL-14; BL) and CD69–PE–Cy7 (H1.2F3; eB). For the determination of thymocyte populations: CD3–APC–Cy7 or –AF700 (17A2; eB), CD4–PE (RM4-4, BL), CD8–APC (53-6.7, eB), CD62L–PerCP–Cy5.5 (MEL-14; BL), CD69–PE–Cy7 (H1.2F3; eB), CD44–AF700 or –PECy7 (IM7; eB), CD25–AF488 (3C7; eB), cKit–PerCP–eFluor710 or –APC–eFluor780 (CD117; 2B8; eB). DN1 cells (ETPs) were defined as CD4⁺ CD8⁺ CD3⁺ CD44⁺ CD25⁺ cKit⁺; DN2 cells were defined as CD4⁺ CD8⁺ CD3⁺ CD44⁺ CD25⁺; DN3 cells were defined as CD4⁺ CD8⁺ CD3⁺ CD44^{lo/–} CD25⁺; DN4 cells were defined as CD4⁺ CD8⁺ CD3⁺ CD44⁺ CD25⁺. For the determination of bone marrow B-cell populations: B220–AF700, cKit–APC–eFluor780, CD19–PE–Cy7, Flt3–PE, CD24–BV421, CD43–AF488, IgM–PerCP–Cy5.5, IgD–APC. Pro-B cells were defined as B220⁺ cKit⁺ CD19⁺ Flt3⁺ CD24⁺ IgM[–]; pre-B cells were defined as B220⁺ cKit⁺ CD19⁺ Flt3⁺ CD24⁺ IgM[–]; immature B cells were defined as B220⁺ cKit⁺ CD19⁺ Flt3⁺ CD24⁺ IgM⁺, and mature B cells were defined as B220⁺ cKit⁺ CD19⁺ Flt3⁺ CD24^{lo/–} IgM^{+/–} IgD⁺. To quantify bone marrow cell populations with or without BrdU^{18–20,35–37}: Lineage (Lin) cocktail–eFluor450 (17A2, RA3-6B2, M1/70, Ter-119, RB6-8C5; eB), Flt3–PE (CD135; A2F10; eB), cKit–PerCP–eFluor710 or –APC–eFluor780 (CD117; 2B8; eB), Sca-1–PE–Cy7 (Ly6A/E; D7; eB), IL-7Rα–APC–eFluor780, –FITC, or BV421 (CD127; A7R34; eB or SB/199; BD), CD34–APC (RAM34; BD), FcγRII/III–AF700 (93; eB), BrdU–PerCP–eFluor710 (BU20A; eB). For low abundance markers (cKit, Sca-1, IL-7Rα, CD34, FcγRII/III, CD25 or CD44), fluorescence minus one (FMO) controls were used to set staining gates.

For SIP₁ staining, tibiae bone marrow were spun into a tube containing 35 μl EDTA, then immediately fixed with 0.1% PFA on ice for 1 h. After washing with 100 nM HEPES in PBS, nonspecific interactions were blocked with anti-CD16/32 at room temperature for 15 min. Cells were then aliquoted for anti-SIP₁ or isotype staining (R&D) overnight at 4 °C. The next day, cells were washed twice with PBS-HEPES and stained to identify bone marrow populations.

Phosphoflow for phospho-Stat5 and phospho-Erk1/2. Whole bone marrow cells were resuspended in Lyse/Fix buffer (BD Biosciences) and processed according to manufacturer's instructions. To permeabilize, 100% methanol pre-cooled at –80 °C was added and cells were vortexed for 15 s before 45 min incubation on ice. After washing with 3% FBS in PBS, cells were incubated in Fc block (anti-CD16/32) and Innovex Biosciences peptide Fc block at room temperature for 20 min before incubation with antibody cocktails. For detection of phospho epitopes in CLPs, the following antibodies were used: Lineage cocktail–PE, cKit–PerCP–eFluor 710, Sca-1–PE–Cy7, and CD127–BV421. Anti-p-Stat5 (pY694) and p-Erk1/2 (pT202/pY204) were conjugated to AF488. As controls, splenocytes were isolated from wild-type mice and incubated with IL-7 (10 ng ml^{–1}; p-Stat5) or PMA (5 ng ml^{–1}; p-Erk1/2) or left unstimulated. After 15 min, Lyse/Fix buffer with 30% FBS was immediately added to cells to preserve the phosphor epitopes. Cells were then processed as described earlier for bone marrow samples. CD11b⁺ cells were considered the negative control for p-Stat5 in response to IL-7 stimulation. CD19⁺ cells were considered the positive controls for p-Stat5 and p-Erk1/2 (refs 34, 38).

Drug administration. FTY720 (fingolimod, trade name Gilenya) was administered orally at a single dosage of 0.5 mg kg^{–1} in 200 μl of 2% 2-hydroxypropyl-β-cyclodextran (HβC; Sigma-Aldrich). AUY954 was dissolved in dimethylsulphoxide (DMSO) and administered orally at a single dosage of 3 mg kg^{–1} in 200 μl of 2% HβC. Both FTY720 and AUY954 were gifts from Novartis Pharma AG. SEW2871 (Cayman Chemical) was dissolved in 1:4 Tween-20 and ethanol, then administered twice at 0 h and 12 h at a dosage of 20 mg kg^{–1} orally in 200 μl of HβC before mice were killed at 24 h. For all studies, control animals were treated with the appropriate vehicle.

Albumin ELISA. Albumin concentrations in blood and lymph plasmas were determined using the mouse albumin ELISA kit from Bethyl Labs, according to

manufacturer's instructions. Blood plasma was diluted $1:1.6 \times 10^6$ and lymph plasma $1:8 \times 10^5$ before analyses.

ApoM and ApoA1 immunoblots. One microlitre of mouse blood plasma was first diluted in 10 μ l PBS, then diluted serially 1:1 in PBS, and 1 μ l of diluted lymph plasma was further diluted in 10 μ l PBS. Samples were then denatured by boiling in SDS sample buffer and separated on a 12% denaturing gel. After transfer to PVDF membrane, blots were incubated overnight at 4 °C with 1:1,000 dilution of antibodies against ApoM (rabbit anti-human; Genetex) or ApoA1 (goat anti-mouse; AbCam) in 5% milk in Tween-20 Tris-buffered saline (TBS-T). Secondary antibodies were incubated at 1:10,000 dilution in 1% TBS-T for 1 h at room temperature. For approximation of ApoM concentrations, bands for the blood plasma dilution were analysed by densitometry using ImageJ software (NIH) to create a standard curve. Undiluted blood plasma ApoM concentrations were assumed to be approximately 900 nM. Densitometric values obtained for ApoM bands from lymph plasma were compared to the standard curve and an approximate concentration calculated, which was then normalized to lymph volume. ApoA1 was used as an internal control.

S1P quantitation. Bioactive sphingolipids were extracted from blood, lymph or HDL preparations by a single-phase ethyl acetate:isopropanol:water process and quantitated by electrospray ionization-tandem mass spectrometry (ESI-MS/MS) as described previously³⁹.

Integrin blockade. Mice were injected intraperitoneally with 100 μ g of anti-integrin α 4 (BD Pharmingen, clone M17/4) and 100 μ g of anti-integrin α L (Millipore, clone PS/2)⁴⁰ in a total of 200 μ l. Fourteen hours later, mice were killed and brachial and inguinal lymph nodes collected. A single-cell suspension was then analysed by flow cytometry.

Administration and quantitation of BrdU. BrdU was solubilized in 10 mM NaOH/H₂O. Four milligrams BrdU (Sigma) in 200 μ l PBS was injected intraperitoneally into each mouse 2.5–3 h before they were killed^{37,41,42}. Control mice were injected with 200 μ l of PBS alone. For quantitation of BrdU incorporation, after isolation and red blood cell (RBC) lysis, cell suspensions were fixed in 4% paraformaldehyde (BD) at room temperature for 10 min, then washed in 5% FBS followed by PBS. Cells were permeabilized in 100 μ l of 0.125% Dawn dishwashing detergent (Procter and Gamble) in 1 \times PBS for 15 min at room temperature, then washed twice with 1% FBS in PBS. Cells were resuspended in DNase buffer (10% FBS in 1 mM CaCl₂ and 0.1 mM MgCl₂) containing 150 U of DNase per sample and incubated at 37 °C for 45 min before washing twice in PBS. Samples were then resuspended in Fc blocking buffer and incubated with antibody cocktail for analysis by flow cytometry.

Cell cycle analysis using Hoechst 33342 DNA incorporation. Freshly isolated cells were stained with antibody cocktail (anti-Lineage, anti-cKit and anti-Sca-1), washed, then resuspended in 10 μ g ml⁻¹ Hoechst 33342 (Life Technologies) in 1 \times PBS and incubated at room temperature for 45 min before fixation with 0.1% PFA and flow cytometric analysis.

Bone marrow IL-7 quantitation. Complete bone marrow was collected in 5 μ l 0.1 M EDTA, then gently triturated in 200 μ l PBS. Cell suspensions were then centrifuged at 2,000g for 5 min at 4 °C, then supernatants were removed to a new tube and stored at -80 °C until analysis. IL-7 expression was quantified by ELISA according to the manufacturer's instructions (R&D).

Multiplex quantitative real-time RT-PCR. B cells were isolated from whole spleen single-cell suspensions using the EasySep mouse CD19-positive selection kit (StemCell Technologies), followed by the EasySep mouse T-cell isolation kit for negative selection of T cells. Bone marrow from wild-type mice was subjected to negative selection for Lin⁻ cells, then LSK (Lin⁻ cKit⁺ Sca-1⁺) cells and CLPs (Lin Kit^{lo} Sca^{lo} IL-7R α ⁺ Flt3⁺) were collected by FACS. RNA was isolated using Trizol and concentrated using the RNA Clean and Concentrator kit (Zymo Research). Real-time RT and TaqMan PCR were performed using the TaqMan RNA-to C_T 1-step kit (Life Technologies), and primers specific for mouse *S1pr1*. Reactions were normalized to expression of the housekeeping gene *Nidogen* within the same tube using the $\Delta\Delta 2^{-C_t}$ method⁴³ and data are represented as fold change compared to expression by splenic B cells.

Ex vivo ApoM staining of bone marrow cells. Bone marrow was spun into a tube containing 35 μ l EDTA and the femurs flushed with PBS and resuspended. Cells were fixed with 0.1% paraformaldehyde at room temperature for 30 min and washed with PBS. Non-specific interactions were first blocked by incubation with Innovex Biosciences peptide Fc block at room temperature for 15 min, then 10% fatty-acid-free bovine serum albumin, Fraction V (FAF-BSA), 5% milk and anti-CD16/32 were added to 1 ml for an additional 10 min. This was used in place of the suggested separation medium when cells were processed for lineage-negative depletion, followed by Sca-1 positive selection (both kits from StemCell Technologies). Cells were then washed twice with BSA/milk by gentle mixing, then returning them to the magnet and pouring off liquid. Two drops of peptide Fc block with anti-CD16/32 were added to each tube and incubated at room

temperature for 5 mins. An aliquot of cells was set aside for staining with secondary antibody alone; the remainder was incubated with 1:200 dilution of rabbit anti-ApoM (GeneTex #62234) for 30 min at room temperature. Cells were washed on the magnetic column three times with BSA/milk before incubation with a 1:5,000 dilution of donkey anti-rabbit AF546 (Life Technologies) for 30 min at room temperature. After washing again on the magnetic column, cells were incubated in PBS containing DAPI before imaging with an Olympus Fluoview Fv10i confocal microscope using the $\times 60$ water objective. Intensities for the blue and red channel were adjusted for the *Apom*^{-/-} cells in accordance with their spectra and, once set, the same settings were used for the wild-type images. Thus, channel intensity correction was the same for each sample and no gamma correction was used.

Ex vivo IL-7Ra staining of bone marrow cells from S1P₁ GFP signalling mice. Bone marrow was spun into a tube containing 35 μ l EDTA and the femurs flushed with, and resuspended in, 0.1% paraformaldehyde. Cells were fixed at room temperature for 12 min and washed with 1% FBS in PBS before being subjected to lineage depletion followed by Sca-1 positive selection. Cells were then stained with either Lin cocktail or biotinylated anti-IL-7R α primary antibody (clone SB/199; BD) followed by streptavidin-AF596 secondary antibody (Jackson Immuno-research). All cells were incubated in DAPI in PBS before imaging by confocal microscopy, as described earlier.

In vitro haematopoiesis and lymphopoiesis. Lineage-negative (Lin⁻) bone marrow cells from wild-type or S5A mice were isolated using a negative selection magnetic bead kit (StemCell Technologies). In some experiments, CLPs from wild-type mice were sorted by FACS. Five-thousand cells were then incubated in 350 μ l methylcellulose medium HPP-SP1 (MC; HemoGenix) containing IL-3, IL-6, stem cell factor (SCF), thrombopoietin (TPO), Flt3L, BSA and 5% FBS and supplemented with 10 ng ml⁻¹ rmIL-7 (Peprotech). Control wells contained only this complete medium. Data from at least two duplicate wells per experimental condition were averaged. Cells from each mouse were assayed individually and not pooled. The stock solution of W146 was prepared in filtered Na₂CO₃ buffer and sonicated until dissolved, then added to MC at a final concentration of 1 μ M. AUY954 was dissolved in DMSO and diluted in MC for a final concentration of 100 nM and 0.001% DMSO. FTY720p was dissolved in 1 \times PBS and added to MC. The vehicle control for the *in vitro* agonist/antagonist studies was 0.001% DMSO. Human and mouse HDLs were isolated by ultracentrifugation of plasma from healthy donors or mice as previously described and added to MC at the concentrations indicated^{44–46}. The content of S1P in HDL preparations was determined by high-performance liquid chromatography (HPLC) and LS/MS/MS as described earlier⁴⁶. For S1P neutralization by anti-S1P antibody, 50 μ g ml⁻¹ HDL was incubated with 25 μ g ml⁻¹ of either anti-S1P antibody (Alfresa Pharma Corp) or mouse IgM (Bethyl Labs) at 37 °C for 1 h, then added to the MC⁴⁷. Albumin-S1P was prepared by reconstituting stock S1P (in ethanol) in 0.1% fatty-acid-free BSA (FAF-BSA; Fisher), then sonicating for 20 s. Dilutions were then made in 0.1% FAF-BSA before adding to MC. All MC incubations of Lin⁻ cells were incubated for 10 days at 37 °C and 5% CO₂.

Flow cytometric analysis of in vitro incubations. For cell collection, 500 μ l of PBS was added to each well, incubated for 10 min, then gently triturated to resuspend the cell culture/MC mixture. Resuspended cells were removed to a 1.5 ml microcentrifuge tube and cells remaining in the well were collected with an additional 500 μ l of PBS. Cells were then centrifuged at 700g for 10 min at 4 °C, washed once with PBS, then processed for incubation with antibody cocktail and flow cytometric analysis as earlier.

Induction of EAE. EAE was induced by subcutaneous injection on both flanks and base of the tail of 200 μ g myelin oligoglycoprotein (MOG_{35–55}) peptide (Alpha Diagnostic Intl.) emulsified in 200 μ l of complete Freund's adjuvant (Difco). Two-hundred nanograms of *Bordetella pertussis* toxin (List Biological Laboratories) was injected intraperitoneally in 200 μ l on the day of immunization and 48 h later^{5,13}. Clinical scores were determined based on the following scale: 0, normal; 1, tail weakness; 2, tail paralysis and hind limb weakness; 3, complete hind limb paralysis; 4, hind limb paralysis with forelimb weakness; and 5, moribund or death¹³. For histopathological analysis of brain tissue sections, brains were removed and fixed in 10% neutral-buffered formalin overnight at room temperature. After washing in PBS, brains were placed in 70% ethanol before paraffin embedding. Tissue sections were then stained with haematoxylin and eosin (HistoServe).

Isolation of CNS-infiltrating immune cells. Brains from wild-type and *Apom*^{-/-} mice were minced, then digested in a mixing incubator for 1 h at 37 °C and 1,300 r.p.m. in a cocktail of Liberase (1 U per 1 ml; Roche) and DNase (1 mg ml⁻¹; Sigma) in Hank's balanced salt solution (HBSS). Tissues were further disaggregated by repeated passage through an 18G needle, then brought to 1 ml with HBSS. Leukocytes were isolated on a discontinuous Percoll gradient according to Pino and Cardona⁴⁸ and analysed by flow cytometry.

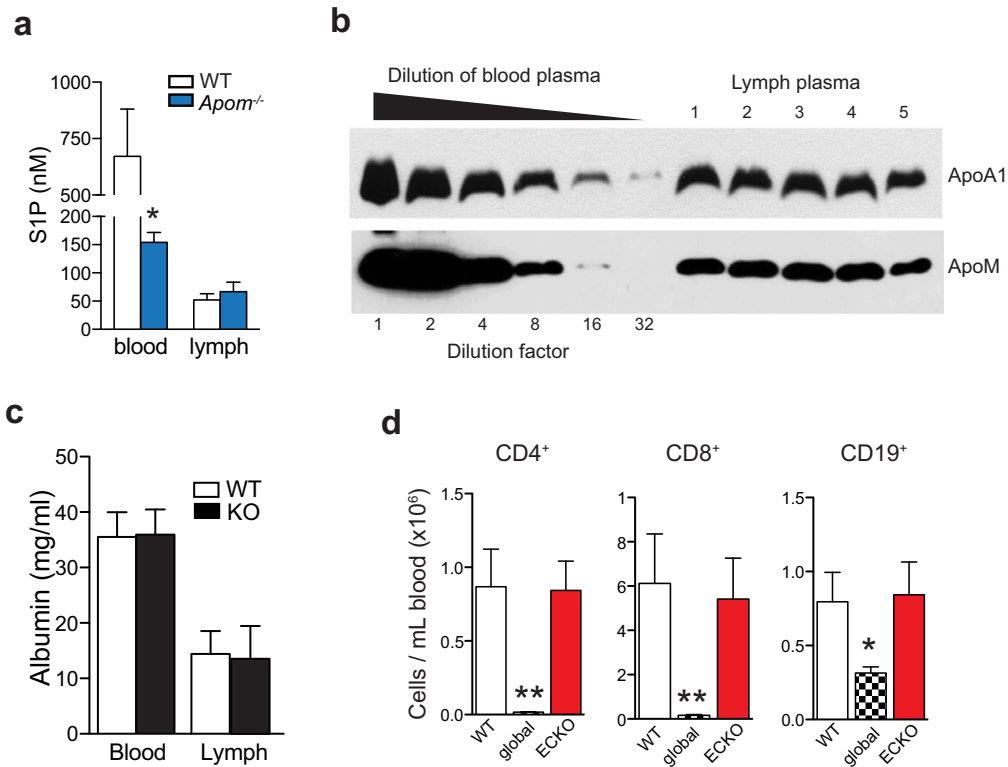
In vitro lymphocyte restimulation. Spleens from wild-type or *Apom*^{-/-} mice were isolated 9 days after MOG_{35–55} peptide immunization. After disaggregation and erythrocyte lysis with ammonium chloride, cells were cultured in 96-well plates at a density of 5×10^5 cells with increasing concentrations of MOG_{35–55} peptide. After 96 h of incubation, numbers of CD4⁺ cells were quantified by flow cytometry¹³.

Evans blue dye extravasation measure of vascular permeability. Pulmonary vascular leakage was measured by Evans blue dye (EBD) accumulation assay. EBD (0.5% in PBS) was injected via tail vein 90 min before tissue harvest. Pulmonary vasculature was perfused with 5 ml PBS through the right ventricle, while allowing the perfusate to drain from an incision in the left ventricle. Lungs were then weighed and dried at 56 °C overnight. Dry lung weight was measured again, and EBD was dissolved in formamide at 37 °C for 24 h and quantitated spectrophotometrically at 620 and 740 nm.

TMR-dextran brain and spinal cord vascular permeability. Eight days after MOG immunization, MOG-immunized mice, mice injected with pertussis toxin alone, or unimmunized mice were injected via tail vein with 100 µl of 25 mg ml⁻¹ TMR-dextran in 1 mM CaCl₂. After 18 h, mice were killed by CO₂ and perfused with approximately 10 ml PBS. Brain and spinal cord were removed and protected from light from this point. After weighing, tissues were homogenized in 1% Triton X-100 in PBS then spun at 4 °C for 25 min at 20,000g. Supernatants were transferred to new tubes until fluorometric analysis at 540/590.

Statistical analyses. Statistical analyses were performed using GraphPad Prism software v.6.0. Two-tailed Student's *t*-test was used for direct comparison of two groups. Analysis of variance (ANOVA) followed by Dunnett's for comparison to control group or Tukey's to compare all groups was used to determine significance between three or more test groups. Differences in EAE clinical scores were determined using the Mann-Whitney U-test for analysis of the area under the curve and at individual time points. For all analyses, α was set to 0.05 and *P* values ≤ 0.05 were considered significantly different.

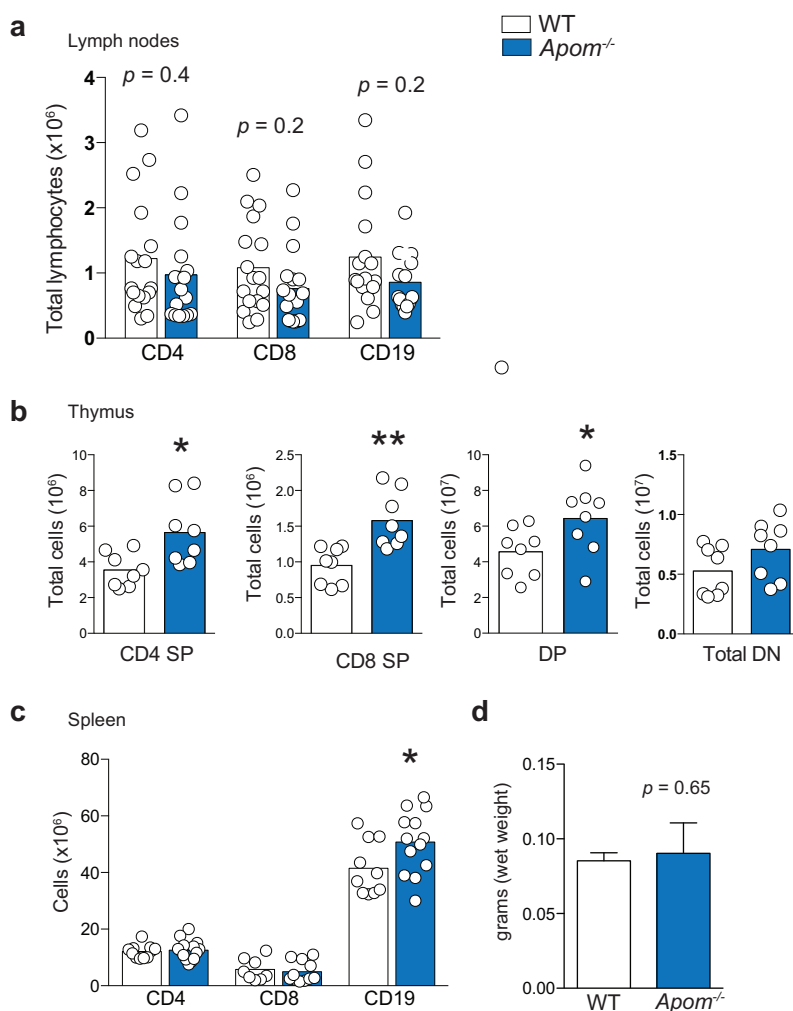
31. Christoffersen, C. *et al.* Effect of apolipoprotein M on high density lipoprotein metabolism and atherosclerosis in low density lipoprotein receptor knock-out mice. *J. Biol. Chem.* **283**, 1839–1847 (2008).
32. Pitulescu, M. E., Schmidt, I., Benedito, R. & Adams, R. H. Inducible gene targeting in the neonatal vasculature and analysis of retinal angiogenesis in mice. *Nature Protocols* **5**, 1518–1534 (2010).
33. Matloubian, M. *et al.* Lymphocyte egress from thymus and peripheral lymphoid organs is dependent on S1P receptor 1. *Nature* **427**, 355–360 (2004).
34. Kalaitzidis, D. & Neel, B. G. Flow-cytometric phosphoprotein analysis reveals agonist and temporal differences in responses of murine hematopoietic stem/progenitor cells. *PLoS ONE* **3**, e3776 (2008).
35. Ding, L. & Morrison, S. J. Hematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* **495**, 231–235 (2013).
36. Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
37. Passegué, E., Wagers, A. J., Giuriato, S., Anderson, W. C. & Weissman, I. L. Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates. *J. Exp. Med.* **202**, 1599–1611 (2005).
38. Krutzik, P. O., Clutter, M. R. & Nolan, G. P. Coordinate analysis of murine immune cell surface markers and intracellular phosphoproteins by flow cytometry. *J. Immunol.* **175**, 2357–2365 (2005).
39. Bielawski, J. *et al.* Comprehensive quantitative analysis of bioactive sphingolipids by high-performance liquid chromatography-tandem mass spectrometry. *Methods Mol. Biol.* **579**, 443–467 (2009).
40. Pham, T. H. M., Okada, T., Matloubian, M., Lo, C. G. & Cyster, J. G. S1P₁ receptor signaling overrides retention mediated by G α_i -coupled receptors to promote T cell egress. *Immunity* **28**, 122–133 (2008).
41. Tough, D. F., Sprent, J. & Stephens, G. L. Measurement of T and B cell turnover with bromodeoxyuridine. *Curr. Protoc. Immunol.* Chapter 4, Unit 4.7 (2007).
42. Wright, D. E. *et al.* Cyclophosphamide/granulocyte colony-stimulating factor causes selective mobilization of bone marrow hematopoietic stem cells into the blood after M phase of the cell cycle. *Blood* **97**, 2278–2285 (2001).
43. Brown, C. R., Blaho, V. A. & Loiacono, C. M. Susceptibility to experimental Lyme arthritis correlates with KC and monocyte chemoattractant protein-1 production in joints and requires neutrophil recruitment via CXCR2. *J. Immunol.* **171**, 893–901 (2003).
44. Hammad, S. M. *et al.* Blood sphingolipidomics in healthy humans: impact of sample collection methodology. *J. Lipid Res.* **51**, 3074–3087 (2010).
45. Hammad, S. M., Al Gadban, M. M., Semler, A. J. & Klein, R. L. Sphingosine 1-phosphate distribution in human plasma: associations with lipid profiles. *J. Lipids* **2012**, 180705 (2012).
46. Min, J. K., Yoo, H. S., Lee, E. Y., Lee, W. J. & Lee, Y. M. Simultaneous quantitative analysis of sphingoid base 1-phosphates in biological samples by o-phthalaldehyde precolumn derivatization after dephosphorylation with alkaline phosphatase. *Anal. Biochem.* **303**, 167–175 (2002).
47. Wolfrum, C., Poy, M. N. & Stoffel, M. Apolipoprotein M is required for pre β -HDL formation and cholesterol efflux to HDL and protects against atherosclerosis. *Nature Med.* **11**, 418–422 (2005).
48. Pino, P. A. & Cardona, A. E. Isolation of brain and spinal cord mononuclear cells using percoll gradients. *J. Vis. Exp.*, (2011).



Extended Data Figure 1 | Blood and lymph ApoM and albumin in wild-type and *Apom*^{-/-} mice and blood cell numbers in *S1pr1* global and ECKO mice.

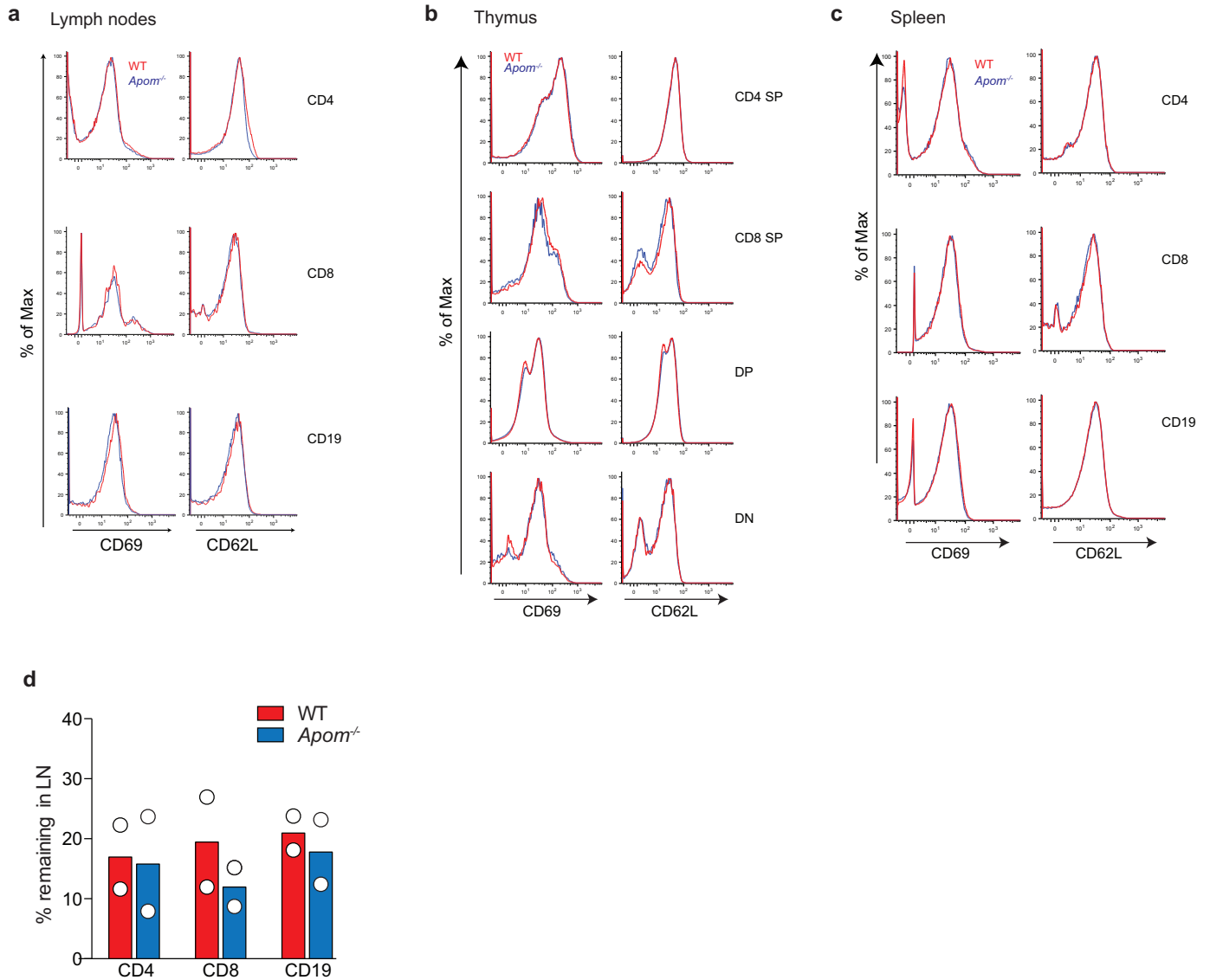
a, Concentrations of blood and lymph S1P were determined by liquid chromatography tandem mass spectrometry (LC-MS/MS). Bars represent means \pm s.d. $n = 4$ as described in Methods. WT, wild type. **b**, Western blots of ApoA1 and ApoM in lymph of wild-type mice. One microlitre of wild-type blood plasma was serially diluted 1:1, and 1 μ l of diluted lymph plasma from 5 animals was analysed for ApoA1 and ApoM protein levels. **c**, Determination by

enzyme-linked immunosorbent assay (ELISA) of albumin concentrations in the blood and lymph of wild-type or *Apom*^{-/-} (KO) mice. Bars are mean \pm s.d. Wild type, $n = 5$; *Apom*^{-/-}, $n = 7$. **d**, Quantification of CD4⁺, CD8⁺ and CD19⁺ cells in the blood of wild-type, *S1pr1*^{fl/fl} *Rosa26-Cre-ER*^{T2} (global) or *S1pr1*^{fl/fl} *Cdh5-Cre-ER*^{T2} (ECKO) mice. Bars are mean \pm s.d. Wild type, $n = 5$; global or ECKO, $n = 6$. * $P < 0.05$, ** $P < 0.005$, as compared to wild-type values.



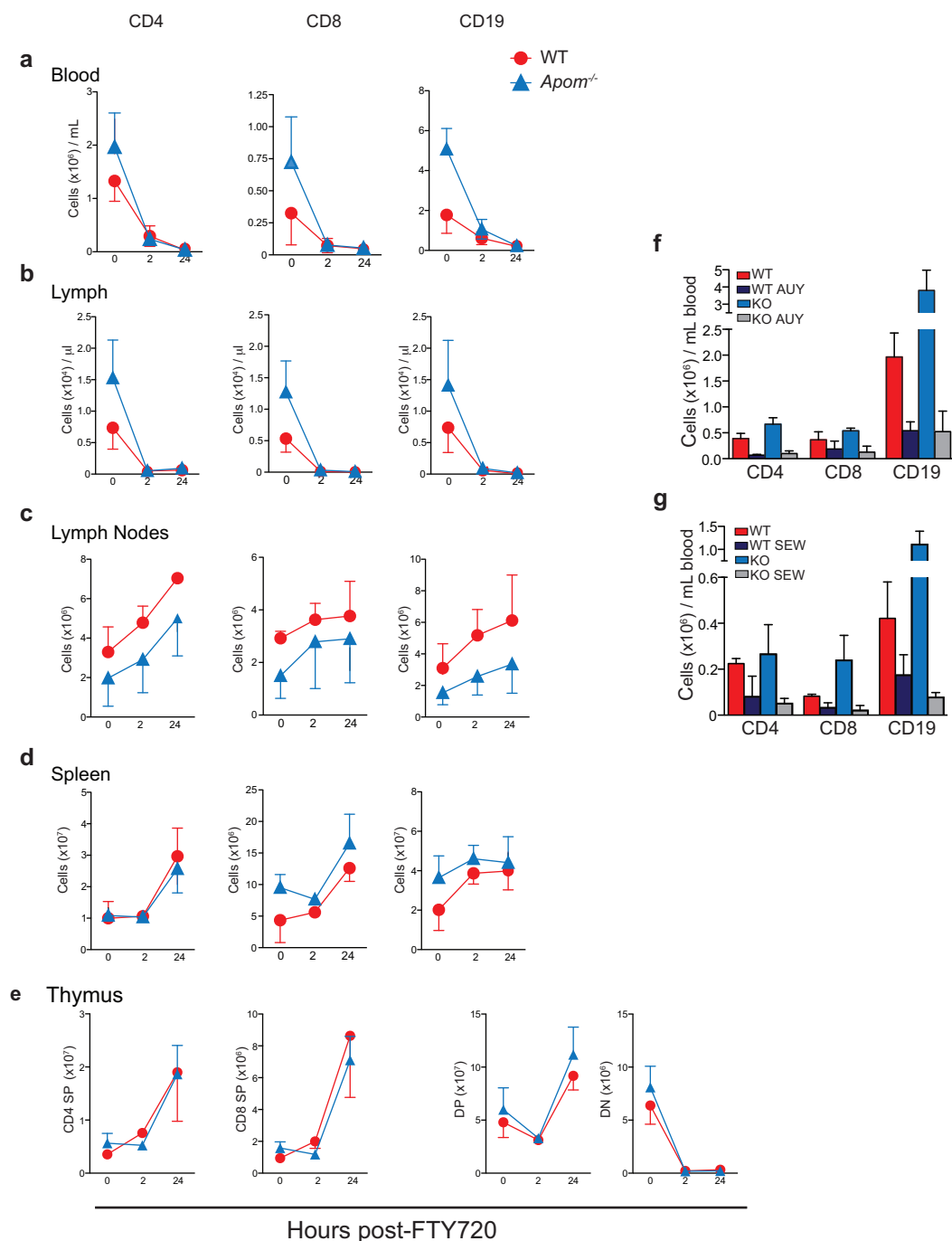
Extended Data Figure 2 | Cell populations of the lymph node, thymus and spleen in *Apom*^{-/-} and wild-type mice. **a**, Quantification of CD4⁺, CD8⁺ and CD19⁺ cells in two brachial and two inguinal lymph nodes combined. Data are compiled from five experiments. **b**, Quantitation of CD4⁺ single-positive (SP), CD8⁺ single-positive, CD4⁺ CD8⁺ double-positive (DP), and total CD4⁻ CD8⁻ double-negative (DN) cells in thymuses of wild-type or

Apom^{-/-} mice. Data are compiled from three experiments. **c**, Quantitation of CD4⁺, CD8⁺ and CD19⁺ cells in spleens of wild-type or *Apom*^{-/-} mice. Data are compiled from three experiments. **a–c**, Bars represent means and circles represent values obtained from individual mice. $*P < 0.05$. **d**, Total spleen weights (grams of wet weight) from wild-type or *Apom*^{-/-} mice. $n = 6$. Bars represent means \pm s.d.



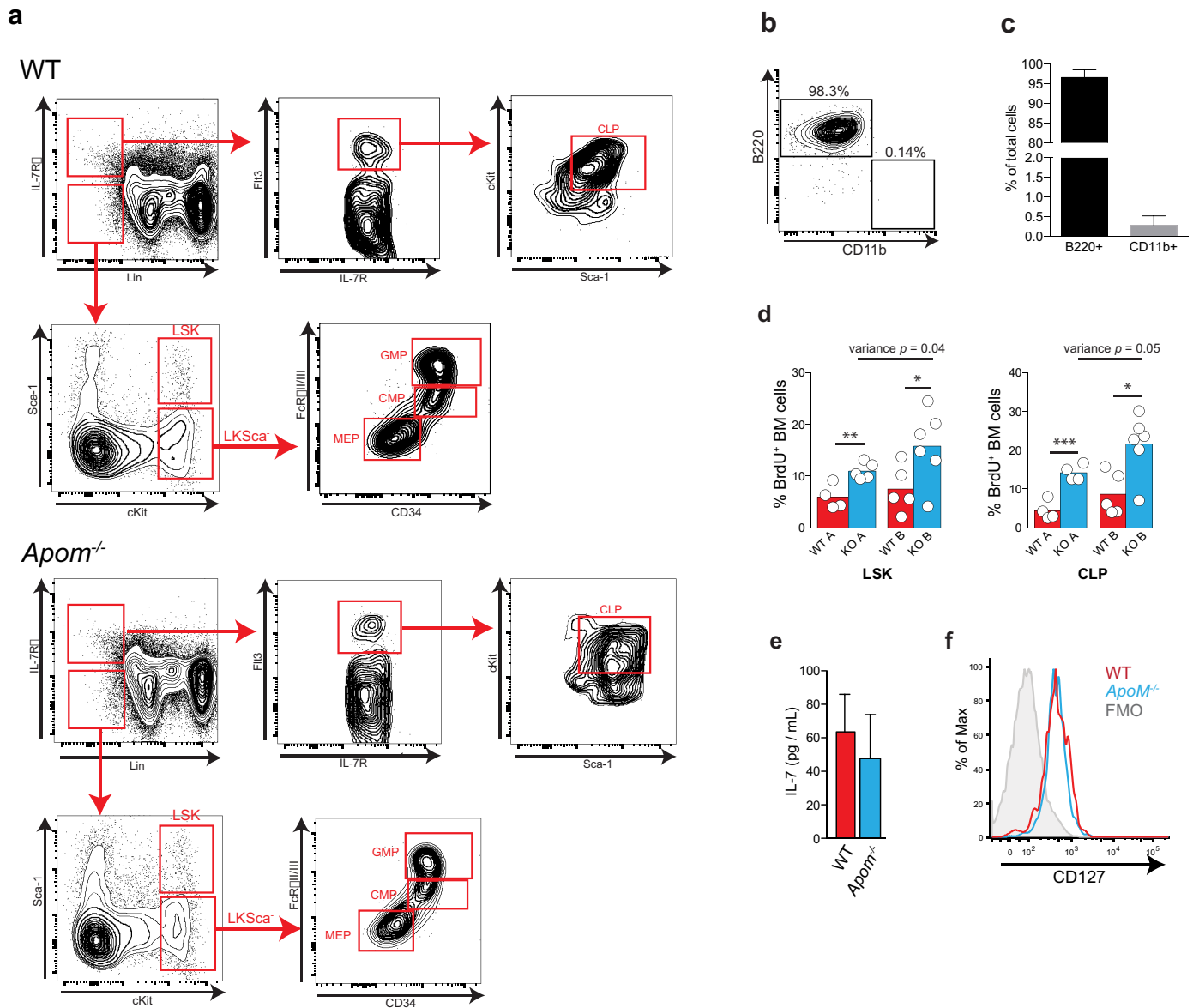
Extended Data Figure 3 | Surface expression of the maturation markers CD62L and CD69 and lymph node egress are unchanged in *Apom^{-/-}* mice. **a–c**, Surface expression of the lymphocyte maturation markers CD62L and CD69 was determined by flow cytometry and representative histograms are shown of staining by lymph node cells (**a**), thymocytes (**b**) or splenocytes

(**c**) quantified in Extended Data Fig. 2 from wild-type (WT; red) and *Apom^{-/-}* (blue) mice. **d**, Percentage of CD4⁺, CD8⁺ and CD19⁺ cells remaining in brachial and inguinal lymph nodes 14 h after administration of alpha-4 and alpha-L integrin-blocking antibodies. Circles represent mean value acquired from three mice in independent experiments; bars represent means.



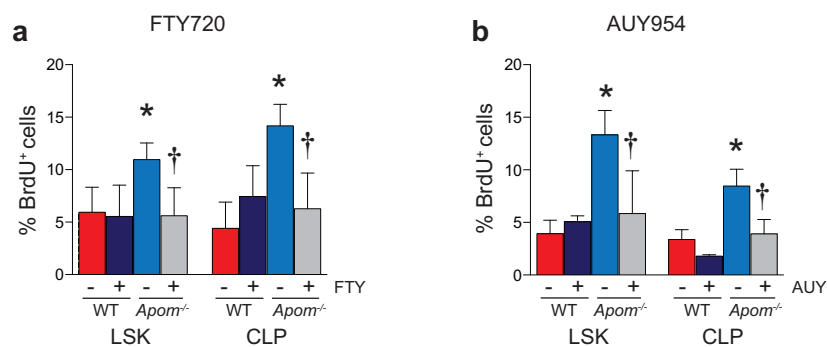
Extended Data Figure 4 | ApoM expression is not required for the lymphopenia response to treatment with FTY720 or the S1P₁-specific agonists AUY954 and SEW2871. a–e, Wild-type (WT; red) and *Apom*^{−/−} (blue) mice were treated with a single dose of FTY720 (0.5 mg kg^{−1} orally), and samples were collected at 0, 2 or 24 h post-treatment. CD4⁺, CD8⁺ and CD19⁺ cells from blood (a), lymph (b), lymph node (c) and spleen (d), and CD4⁺ single-positive, CD8⁺ single-positive, CD4⁺ CD8⁺ double-positive and CD4[−] CD8[−] double-negative cells from the thymus (e) were quantitated by flow

cytometry. Symbols represent means \pm s.d. and graphs are of data compiled from two experiments. 0 h, $n = 5$; 2 h, $n = 6$; 24 h, $n = 4$. f, g, Wild-type and *Apom*^{−/−} (KO) mice were treated with AUY954 (AUY; 1 mg kg^{−1}) (f), SEW2871 (SEW; 20 mg kg^{−1}) (g), or respective vehicle controls. Twenty-four hours after treatment, CD4⁺, CD8⁺ and CD19⁺ cells in the blood were quantitated by flow cytometry. Bars represent means \pm s.d. $n = 4$ for all treatment groups and data are representative of two experiments.



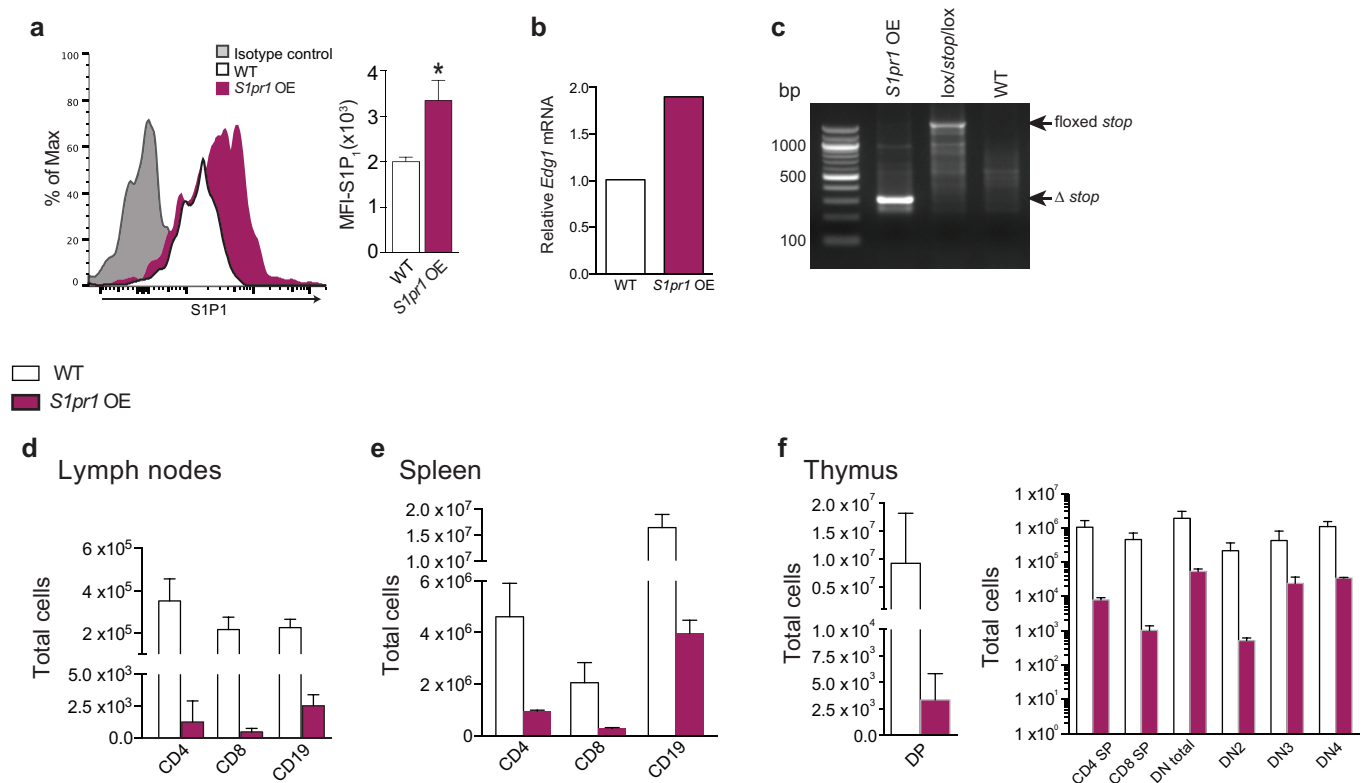
Extended Data Figure 5 | Multiparameter flow cytometry gating scheme for determination of CLPs and other bone marrow populations. **a**, Cells are first gated as Lin⁻ and IL-7R α ⁺. Cells are then gated by Fli3 (Flk2/CD135) versus IL-7R α expression, then further gated by cKit versus Sca-1 expression to define CLPs. Representative staining examples of wild-type (WT; top) or *Apom*^{-/-} (bottom) bone marrow cells are shown. **b**, **c**, Representative flow cytometric plot (**b**) and quantitation (**c**) of B220⁺ CD11b⁻ and CD11b⁺ B220⁻ cells generated from CLPs *in vitro*. CLPs, gated according to the gating hierarchy shown in **a**, were sorted from the bone marrow of wild-type mice and incubated in methylcellulose medium containing growth factors to support both lymphoid and myeloid lineage development. Cells were analysed for B220 and

CD11b expression after 12 days of culture. $n = 6$. **d**, Per cent BrdU incorporation by LSK cells (left) or CLPs (right) in bone marrow of wild-type (red) or *Apom*^{-/-} (blue) mice in two independent experiments (A and B). * $P < 0.05$, ** $P = 0.006$, *** $P = 0.0009$ versus wild type. Equality of variance was determined by an F test. Bars represent means and circles represent values obtained from individual mice. **e**, IL-7 protein in bone marrow supernatants from wild-type and *Apom*^{-/-} mice was quantified by ELISA. $n = 6$ combined from two studies. Bars represent mean \pm s.d. **f**, Representative histogram of IL-7R α (CD127) expression on the surface of wild-type (red) or *Apom*^{-/-} (blue) CLPs. The fluorescence minus one (FMO) control is represented by the grey shaded histogram.



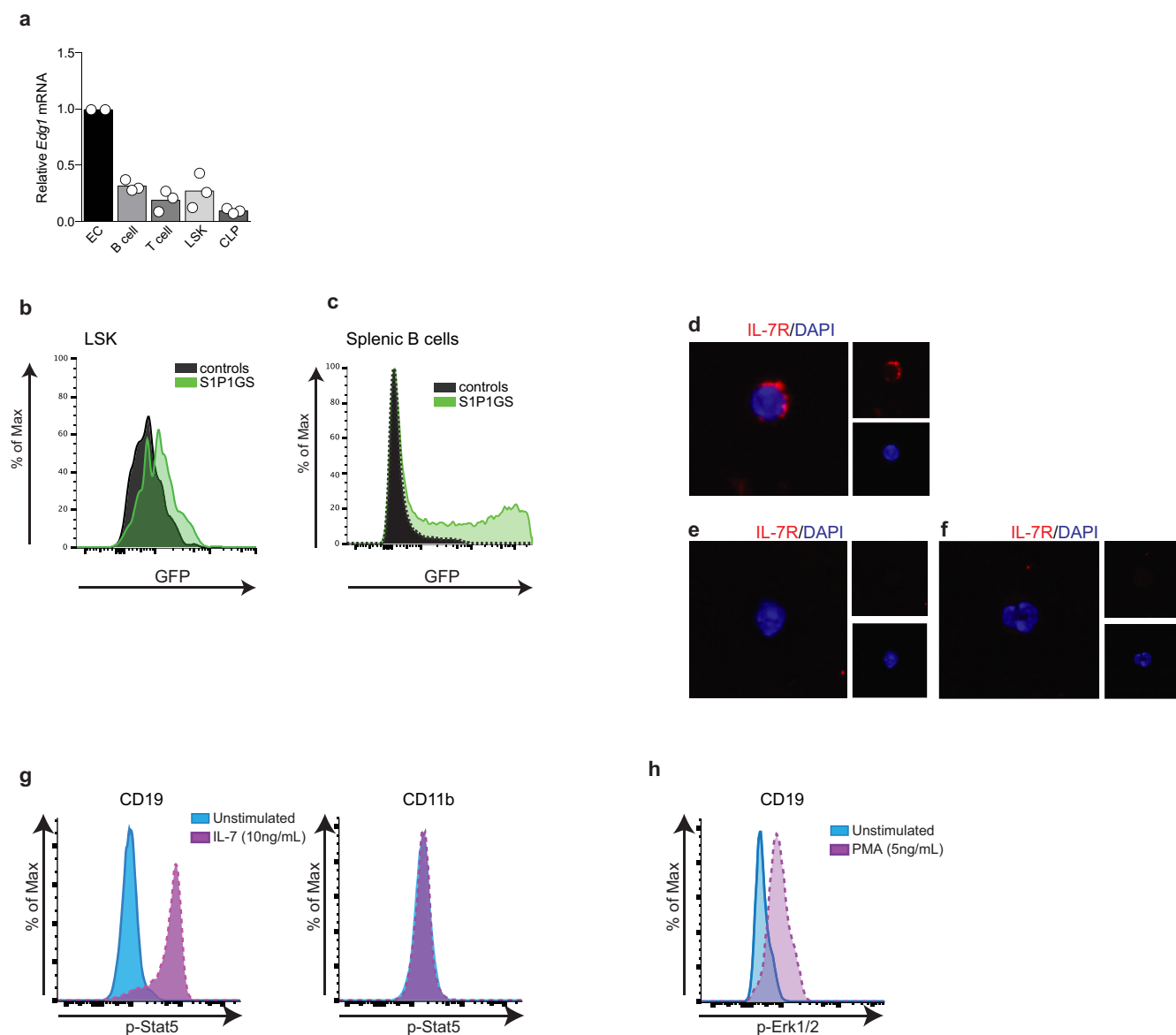
Extended Data Figure 6 | Treatment with the S1P₁ modulator FTY720 or the S1P₁-specific agonist AUY954 suppresses BrdU incorporation by LSK and CLP cells in *Apom*^{-/-} bone marrow. a, b, Per cent BrdU incorporation 24 h after treatment with 0.5 mg kg⁻¹ FTY720 (a) or 1.0 mg kg⁻¹ AUY954 (b). a, Wild type, *n* = 4; *Apom*^{-/-}, *n* = 5 for vehicle-treated and wild type, *n* =

7; *Apom*^{-/-}, *n* = 8 for FTY720-treated groups. b, Vehicle-treated, *n* = 3; AUY954-treated, *n* = 6. Bars represent means ± s.d. and data are compiled from two experiments. **P* < 0.05 versus wild type; †*P* < 0.05 versus vehicle-treated control.



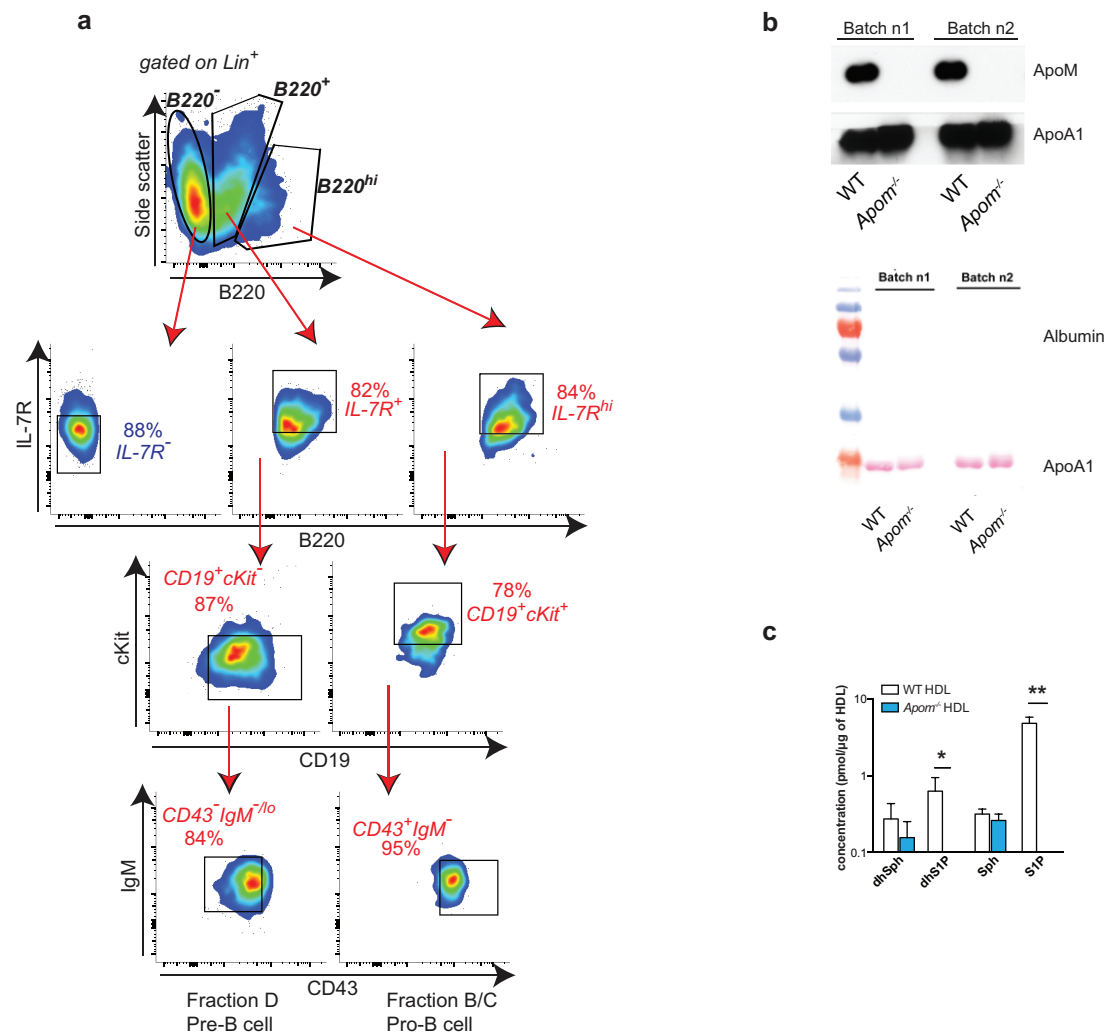
Extended Data Figure 7 | Overexpression of S1P₁ results in marked decreases in lymphocyte populations in the thymus and secondary lymphoid organs. **a**, Representative flow cytometry plots and quantitative MFI of S1P₁ expression by Lin⁺ cells from *S1pr1* OE and wild-type (WT) littermates. *n* = 3. **b**, Relative expression levels of *S1pr1* mRNA in bone marrow cells of *S1pr1* OE mice relative to wild-type mice, as determined by multiplex qRT-PCR. **c**, Representative agarose gel of Cre activation and excision of the floxed stop cassette as assessed by PCR of genomic DNA from bone marrow cells of *S1pr1* OE mice, *lox/stop/lox* littermate, or wild-type littermate. Arrows indicate

amplified DNA fragments corresponding to undeleted or deleted segments **d–f**, Four days after the final dose of tamoxifen, CD4⁺, CD8⁺ and CD19⁺ cells in brachial and inguinal lymph nodes (**d**) and spleen (**e**), and thymic CD4⁺ CD8⁺ double-positive (DP), CD4⁺ and CD8⁺ single-positive (CD4 or CD8 SP), total double-negative (DN), and double-negative subpopulations DN2, DN3 and DN4 (**f**) were quantified by flow cytometry in *S1pr1* OE mice and wild-type littermates. Bars represent means \pm s.d.; *S1pr1* OE, *n* = 6; wild type, *n* = 7 and data are compiled from two experiments.



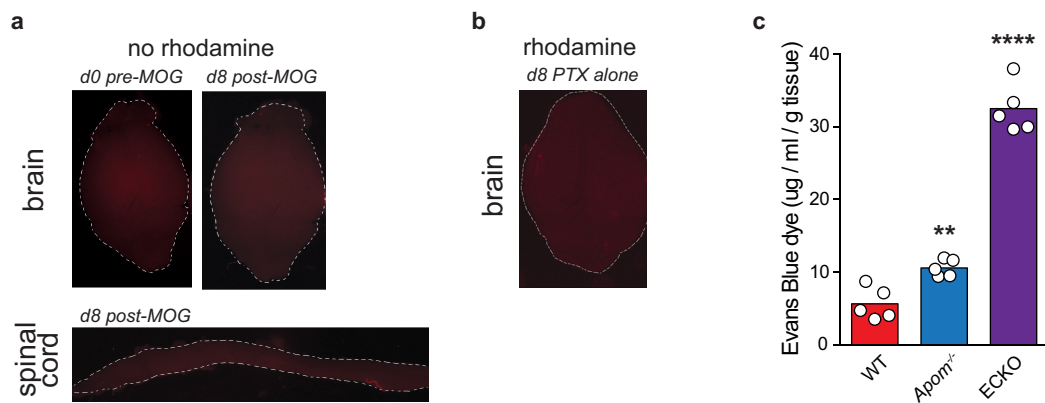
Extended Data Figure 8 | GFP expression by bone marrow cells and splenocytes of S1P₁ GFP signalling mice and stimulation of Stat5 or Erk1/2 phosphorylation in splenocytes. **a**, Relative expression levels of *S1pr1* mRNA in LSK and CLP cells from bone marrow, or splenic B or T cells of wild-type mice relative to mouse endothelial cells, as determined by multiplex qRT-PCR. Bars represent means and circles represent values obtained from individual mice. **b**, Representative histograms of GFP expression by LSK cells from bone marrow of S1P₁ GFP signalling mice (S1P1GS; green) or control (black) mice. **c**, Representative histogram of GFP expression by splenic B cells (CD19⁺) from S1P₁ GFP signalling mice (green) or control (black) mice, demonstrating high *in vivo* GFP expression. **d**, Representative immunofluorescence image of IL-7Rα⁺ cell with CLP morphology from bone marrow of littermate control of S1P₁ GFP signalling mice. Cells were subjected to the same selection process

before immunofluorescence staining. **d–f**, IL-7Rα, red; blue, DAPI. **e, f**, Staining of bone marrow cells from S1P₁ GFP signalling mouse littermate controls demonstrates IL-7Rα staining specificity: cells with CLP morphology stained with secondary alone (**e**) or myeloid cell morphology from bone marrow stained with anti-IL-7Rα (**f**) exhibit no IL-7Rα positivity. **g**, p-Stat5 staining after *in vitro* stimulation of wild-type splenocytes with IL-7 (10 ng ml⁻¹) for 15 min. CD19⁺ cells serve as positive controls for p-Stat5 staining. CD11b⁺ cells serve as negative controls, since they do not have IL-7R and therefore do not respond to IL-7 stimulation with Stat5 phosphorylation. **h**, p-Erk1/2 staining after *in vitro* stimulation of wild-type splenocytes with PMA (5 ng ml⁻¹) for 15 min. CD19⁺ cells serve as positive controls for p-Erk1/2 staining.



Extended Data Figure 9 | *In vitro* lymphopoiesis in the presence of ApoM⁺ HDL generates B cells at different stages of development. **a**, Phenotyping of B220⁺ cell populations generated from wild-type (WT) Lin⁻ bone marrow after 8 days of culture in methylcellulose medium. Initially, cells were gated as B220⁻, B220⁺, or B220^{hi}. From populations expressing B220, pro-B cell (Hardy fractions B/C) and pre-B cell (Hardy fraction D) equivalents were identified. Pro-B cells were defined as IL-7R^{hi}, CD19⁺, cKit⁺, CD43⁺ and

IgM⁻. Pre-B cells were defined as IL-7R⁺, CD19⁺, cKit⁻, CD43⁻ and IgM^{-/lo}. **b**, Western blot analysis of two batches of mouse HDL (batches n1 and n2) isolated from wild-type or *Apom*^{-/-} mice showing the absence of contaminating albumin. **c**, LC-MS/MS analysis of HDL isolated from pooled plasmas of wild-type or *Apom*^{-/-} mice. **P* < 0.05, ***P* < 0.005. *n* = 3. dhSph, dihydrosphingosine; dhS1P, dihydro-sphingosine-1-phosphate; Sph, sphingosine.



Extended Data Figure 10 | Background fluorescence and vascular permeability analysis in brain, spinal cord, and lung. **a**, Representative photographs of auto-fluorescence in whole brains or spinal cord in wild-type mice at day (d)0 or day 8 after MOG₃₅₋₅₅ immunization. 'No rhodamine' denotes tissues from animals that were not injected with TMR-dextran. **b**, Representative photograph of whole brain 16 h after TMR injection of a

mouse injected with pertussis toxin (PTX) alone, without accompanying MOG₃₅₋₅₅ immunization. **c**, Pulmonary vascular permeability as determined by Evans blue dye extravasation in wild-type (WT), *Apom^{-/-}* and S1P₁ ECKO mice. Bars represent means and circles represent values from individual animals. ** $P < 0.005$, **** $P < 0.0001$, as compared to wild-type controls.

Conversion of abiraterone to D4A drives anti-tumour activity in prostate cancer

Zhenfei Li¹, Andrew C. Bishop¹, Mohammad Alyamani¹, Jorge A. Garcia^{2,3}, Robert Dreicer^{2,3}, Dustin Bunch⁴, Jiayan Liu⁵, Sunil K. Upadhyay⁵, Richard J. Auchus^{5,6} & Nima Sharifi^{1,2,3}

Prostate cancer resistance to castration occurs because tumours acquire the metabolic capability of converting precursor steroids to 5 α -dihydrotestosterone (DHT), promoting signalling by the androgen receptor and the development of castration-resistant prostate cancer^{1–3}. Essential for resistance, DHT synthesis from adrenal precursor steroids or possibly from *de novo* synthesis from cholesterol commonly requires enzymatic reactions by 3 β -hydroxysteroid dehydrogenase (3 β HSD), steroid-5 α -reductase (SRD5A) and 17 β -hydroxysteroid dehydrogenase (17 β HSD) isoenzymes^{4,5}. Abiraterone, a steroidal 17 α -hydroxylase/17,20-lyase (CYP17A1) inhibitor, blocks this synthetic process and prolongs survival^{6,7}. We hypothesized that abiraterone is converted by an enzyme to the more active Δ^4 -abiraterone (D4A), which blocks multiple steroidogenic enzymes and antagonizes the androgen receptor, providing an additional explanation for abiraterone's clinical activity. Here we show that abiraterone is converted to D4A in mice and patients with prostate cancer. D4A inhibits CYP17A1, 3 β HSD and SRD5A, which are required for DHT synthesis. Furthermore, competitive androgen receptor antagonism by D4A is comparable to the potent antagonist enzalutamide. D4A also has more potent anti-tumour activity against xenograft tumours than abiraterone. Our findings suggest an additional explanation—conversion to a more active agent—for abiraterone's survival extension. We propose that direct treatment with D4A would be more clinically effective than abiraterone treatment.

The central role and critical requirement for androgen metabolism and androgen receptor (AR) in castration-resistant prostate cancer (CRPC) are demonstrated by the clinical benefit and overall survival benefit conferred by abiraterone (Abi)^{6,7}, which blocks CYP17A1, an enzyme required for androgen synthesis, and enzalutamide, which potently and competitively blocks the AR^{8,9}. Abi (administered in its acetate form for bioavailability) is a steroidal compound and is therefore potentially subject to conversion by steroid-metabolizing enzymes. We hypothesized that the Δ^5 , 3 β -hydroxyl-structure of Abi, which is also present in the natural steroid substrates dehydroepiandrosterone (DHEA) and Δ^5 -androstenediol (A5diol), makes it susceptible to one enzyme conversion by 3 β HSD isoenzymes to its Δ^4 , 3-keto congener (Δ^4 -abiraterone or D4A). This, in turn, would make the steroid A and B rings identical to testosterone, enabling inhibitory interactions with AR and additional steroidogenic enzymes, including SRD5A, which are required for DHT synthesis (Fig. 1a). Such a conversion in peripheral tissues would allow D4A to engage with multiple targets to potentiate its effects on the androgen pathway, providing an alternative explanation for the clinical efficacy of Abi therapy and thus the possibility that direct treatment might be more efficacious.

We found that D4A is detectable in the sera of mice administered Abi acetate (Fig. 1b), as well as in sera from patients with CRPC who

were undergoing treatment with Abi acetate (Fig. 1c, d and Extended Data Fig. 1). In the LAPC4 prostate cancer cell line, which usually has low 3 β HSD activity³, conversion of Abi to D4A is detectable only if 3 β HSD is overexpressed (Fig. 1e and Extended Data Fig. 2a, b). Other tissues such as the mouse adrenal (but not mouse prostate) that have robust endogenous 3 β HSD enzymatic activity also convert Abi to D4A (Extended Data Fig. 2c). These results suggest that D4A is a major metabolite of Abi, requires 3 β HSD for conversion and may confer effects on the tumour that are indirectly due to Abi.

D4A may impinge on multiple steps in the androgen pathway, including CYP17A1, 3 β HSD, SRD5A and direct interaction with AR (Fig. 2a). Although augmented Abi drug exposure may block 3 β HSD,

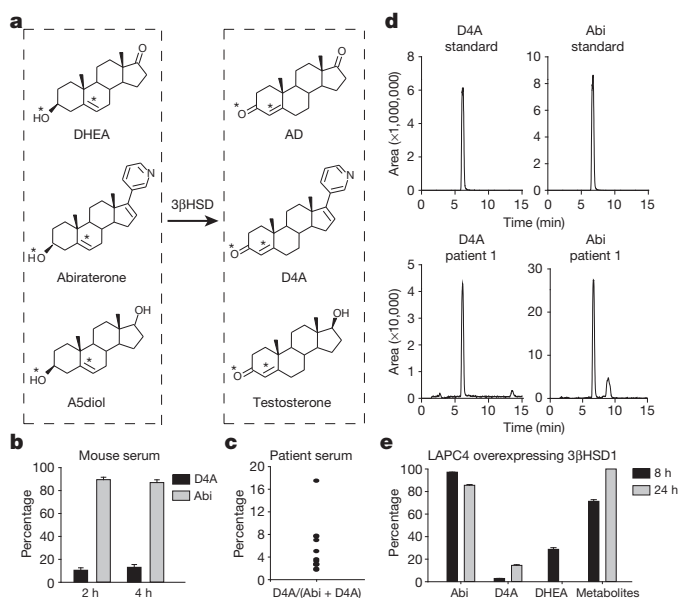


Figure 1 | Structural consequences of the conversion from Abi to D4A that occurs both in mice and in patients, and requires 3 β HSD. **a**, Schematic of Abi conversion to D4A. *Double bond and C3-position for substrates and products of 3 β HSD. **b**, Abi is converted to D4A *in vivo*. Abi acetate was injected intraperitoneally in five mice. Blood was collected 2 h and 4 h after injection. Serum concentrations of Abi and D4A were quantified by mass spectrometry and are represented as the percentage of the sum total of Abi + D4A. **c**, D4A is detectable in all patients ($n = 12$) with CRPC treated with Abi acetate. **d**, Representative mass spectrometry tracing of D4A and Abi from the serum of a patient treated with Abi acetate. **e**, 3 β HSD1 is capable of converting Abi to D4A. LAPC4 cells overexpressing 3 β HSD1 were treated with 10 μ M Abi (or [3 H]DHEA) for 24 h. Abi, D4A (as a percentage of Abi + D4A), [3 H]DHEA and [3 H]DHEA metabolites (labelled metabolites) were separated by HPLC. Results are shown as mean ($n = 3$) \pm s.d. with biological replicates. The experiment was repeated independently at least three times.

¹Department of Cancer Biology, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA. ²Department of Hematology and Oncology, Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA. ³Department of Urology, Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA. ⁴Department of Laboratory Medicine, Pathology and Laboratory Medicine Institute, Cleveland Clinic, Cleveland, Ohio 44195, USA. ⁵Division of Endocrinology and Metabolism, Department of Internal Medicine, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. ⁶Department of Pharmacology, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA.

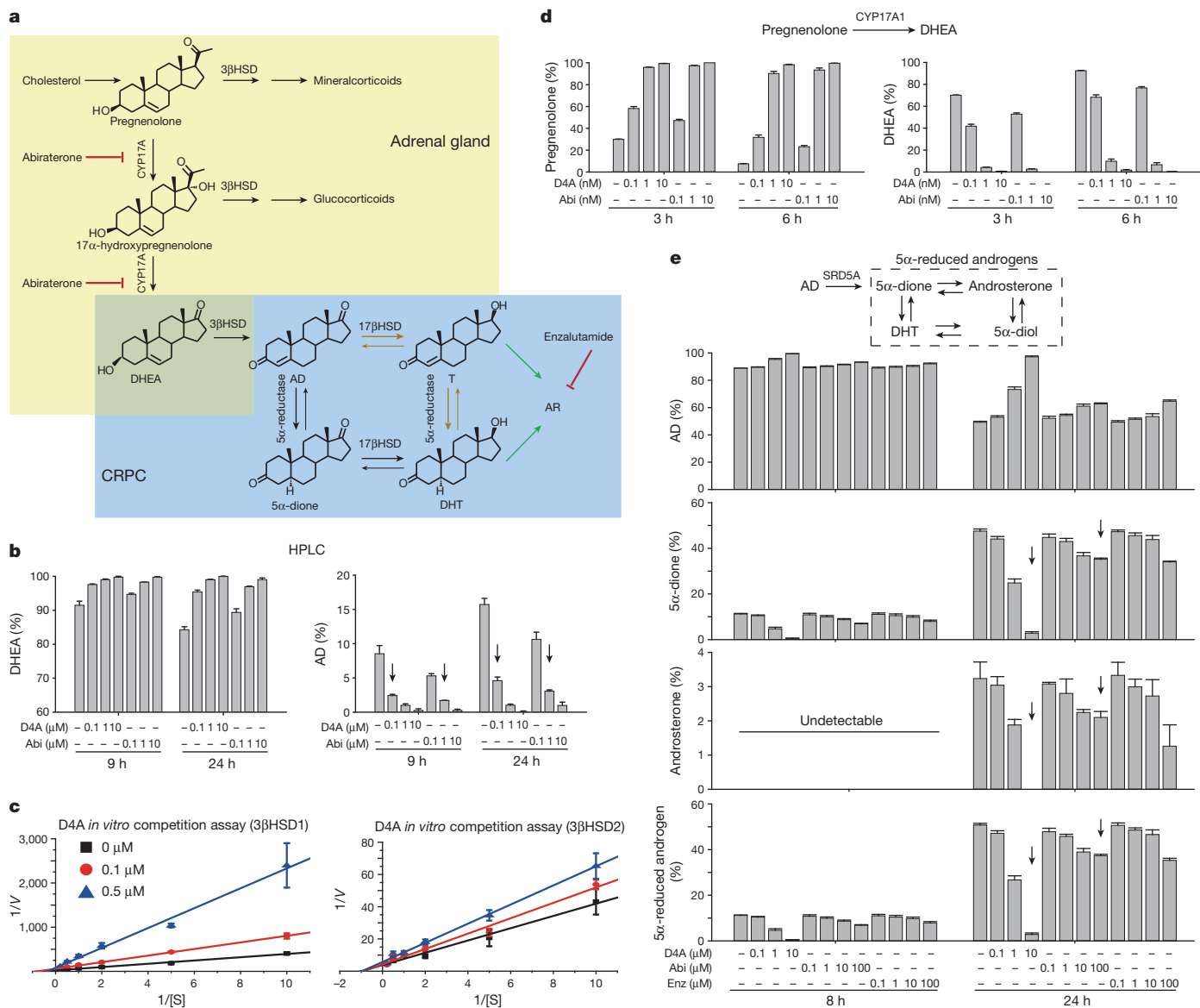


Figure 2 | D4A inhibits 3βHSD, CYP17A1 and 5α-reductase enzymatic activity in the androgen pathway. **a**, Schematic of the steroidogenesis pathway from adrenal precursors to DHT and AR stimulation in CRPC. **b**, D4A inhibits 3βHSD1 activity in LNCaP. Cells were treated with [³H]DHEA with 0.1, 1 or 10 μM D4A and Abi, for 9 and 24 h. DHEA and AD were quantified by HPLC. Arrows denote AD percentage for 0.1 μM D4A and 1 μM Abi treatment groups. **c**, Lineweaver-Burk plots of pregnenolone metabolism and D4A inhibition of human 3βHSD1 and 3βHSD2 activity. D4A shows mixed competitive-non-competitive inhibition for 3βHSD1 and non-competitive inhibition for

3βHSD2. **d**, CYP17A1 inhibition with D4A is comparable to Abi. HEK 293 cells stably expressing CYP17A1 were treated with 0.1, 1 or 10 nM D4A, or Abi, together with [³H]pregnenolone for 3 and 6 h. Pregnenolone and DHEA were separated and quantified by HPLC. **e**, D4A but not Abi or enzalutamide inhibits 5α-reductase activity. LAPC4 cells were treated with [³H]AD and the indicated drug concentrations. Steroids were quantified by HPLC. Arrows denote 5α-reduced androgen percentage for 10 μM D4A and 100 μM Abi treatment groups. All experiments were performed with biological replicates ($n = 3$) and repeated independently three times. All results are shown as mean \pm s.d.

normal dosing probably does not¹⁰. On the other hand, D4A is approximately tenfold more potent than Abi at blocking the conversion of [³H]DHEA by 3βHSD to Δ⁴-androstenedione (AD) in LNCaP and VCaP cells, as assessed by thin-layer chromatography (TLC; Extended Data Fig. 3a) and high-performance liquid chromatography (HPLC; Fig. 2b and Extended Data Fig. 3b). D4A inhibits both human isoenzymes, 3βHSD1 and 3βHSD2, with mixed inhibition kinetics (Fig. 2c). CYP17A1 inhibition is the major direct mechanism of action for Abi¹¹. Structural studies of modified steroidal azoles suggest that the A-ring conformation of D4A does not significantly perturb binding to CYP17A1 (ref. 12). D4A and Abi similarly block conversion of [³H]pregnenolone by CYP17A1 to DHEA (percentage conversion to DHEA after 3 h incubation for vehicle, 1 nM D4A and 1 nM Abi was

70.1%, 4.2% and 2.6%, respectively) by HPLC in intact 293 cells expressing CYP17A1 (Fig. 2d). The Δ⁴, 3-keto-structure of D4A is identical to physiological SRD5A substrates, such as testosterone and AD (Fig. 1a)¹³. To determine the effect of D4A on endogenously expressed SRD5A, LAPC4 cells, which exhibit robust SRD5A enzymatic activity¹³, were treated with D4A, Abi or enzalutamide and cultured in the presence of [³H]AD (the preferred natural substrate of SRD5A1 (ref. 13)). D4A (10 μM) nearly completely blocked conversion from AD to 5α-androstenedione and other 5α-reduced androgens, whereas Abi and enzalutamide had no detectable effect even at a concentration of 100 μM (Fig. 2e).

Abi has been reported to have modest affinity for AR, particularly in the presence of mutations in the ligand-binding domain¹⁴. Conversion

of Abi by 3 β HSD to D4A would provide a 3-keto structure, which is shared by both testosterone and DHT, steroids with the highest affinity for AR (Fig. 1a). To determine how conversion from Abi to the 3-keto structure of D4A affects drug affinity for AR, we performed a competition assay. The affinity of D4A for mutant (expressed in LNCaP, half-maximum inhibitory concentration (IC_{50}) = 5.3 nM) and wild-type (expressed in LAPC4, IC_{50} = 7.9 nM) AR is greater than that of Abi (IC_{50} = 418 and >500 nM, respectively), comparable to or slightly greater than that of enzalutamide (IC_{50} = 24 and 23 nM, respectively; Fig. 3a, b), and clearly greater than bicalutamide, which was the most potent competitive non-steroidal AR antagonist before the introduction of enzalutamide (Extended Data Fig. 4a, b). The affinity of D4A for the AR ligand-binding domain translates to inhibition of DHT-induced AR chromatin occupancy on the *PSA*, *TMPRSS2* and *FKBP5* regulatory elements on chromatin, which is superior to Abi (Extended Data Fig. 4c) and somewhat lower than enzalutamide (Fig. 3c). The incongruity between AR affinity and effects on chromatin occupancy for D4A and enzalutamide is consistent with combined AR antagonism and chromatin binding in an inactive complex, as previously reported for some AR antagonists¹⁵.

We next examined the cumulative results of the effects of D4A on androgen-responsive gene expression. Compared with Abi, D4A clearly better suppresses *PSA*, *TMPRSS2* and *FKBP5* expression induced by DHT, DHEA and R1881 in LNCaP, LAPC4 and C4-2 cell lines (Fig. 3d and Extended Data Fig. 5a, c). D4A inhibits AR target gene expression in a dose-dependent manner (Extended Data Fig. 5b, d).

Comparisons of D4A with enzalutamide on DHT-induced endogenous *PSA* expression demonstrate that D4A is equivalent to enzalutamide against mutant and wild-type AR (Fig. 3e, f and Extended Data Fig. 5c, d). Downstream of androgen-responsive gene expression, effects of D4A and enzalutamide on DHT-stimulated cell growth are equivalent (Fig. 3g), both of which are more potent than Abi.

To determine whether the observed effects of D4A on inhibition of steroid synthesis demonstrated in tissue culture also occur in tumours, effects in two prostate cancer xenograft models with robust 3 β HSD enzymatic activity³ were assessed. Subcutaneous mouse xenograft tumours of VCaP and LNCaP cells, which both harbour a mutant gene encoding a missense in 3 β HSD1 that effectively increases enzyme activity³, were grown in male NSG mice. Fresh tumours were harvested, minced and incubated with [³H]DHEA plus Abi or D4A (0.1–10 μ M). Similar to effects shown in Fig. 2, D4A is tenfold more potent than Abi in blocking conversion from DHEA by 3 β HSD to AD in LNCaP (Fig. 4a) and VCaP xenografts (Fig. 4b). For example, 0.1 μ M D4A is equivalent to 1 μ M Abi for blocking AD accumulation at 48 h in both LNCaP and VCaP xenografts. To test whether the combined effects of D4A on inhibition of steroid synthesis and direct blockade of AR lead to augmented anti-tumour activity compared with Abi, VCaP xenografts were grown subcutaneously in orchiectomized mice with DHEA pellet implantation (to mimic human adrenal physiology; Fig. 4c). Time from initiation of treatment with D4A, Abi acetate or vehicle to tumour progression (>20% increase in tumour volume) was assessed by generating Kaplan–Meier survival curves

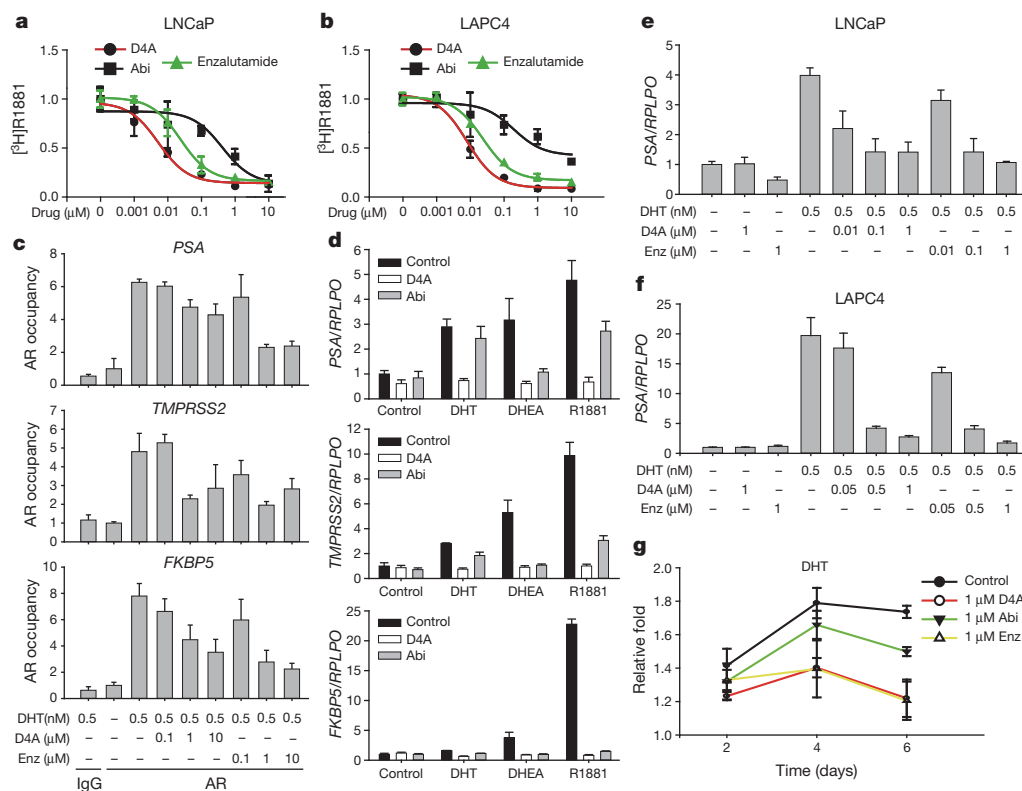


Figure 3 | D4A binds to AR, inhibits AR chromatin occupancy, expression of AR-responsive genes and cell growth. **a, b**, D4A potently binds to both mutant and wild-type AR. D4A, Abi and enzalutamide (Enz) (0.001–10 μ M) were used to compete with 0.1 nM [³H]R1881 for mutated AR (LNCaP) or wild-type AR (LAPC4). Intracellular radioactivity was normalized to protein concentration. **c**, Dose-dependence of D4A versus Enz for inhibition of AR chromatin occupancy. LNCaP cells were treated with the indicated concentrations of DHT, D4A and Enz for 3 h. AR chromatin occupancy for *PSA*, *TMPRSS2* and *FKBP5* was detected with ChIP. AR ChIP was normalized to untreated control for each gene. **d**, D4A inhibits *PSA*, *FKBP5* and *TMPRSS2*

expression. LNCaP cells were treated with DHT (0.5 nM), DHEA (40 nM) or R1881 (0.1 nM) with or without Abi or D4A (1 μ M) for 24 h. Gene expression was detected by quantitative PCR (qPCR) and normalized to *RPLP0*. **e, f**, D4A inhibition of DHT induced *PSA* expression is comparable to Enz in LNCaP and LAPC4. **g**, D4A inhibits DHT (0.5 nM) induced cell growth in LNCaP. Cells were quantified at the indicated time points by assaying DNA content after 2, 4 and 6 days of treatment. Experiments in **a, b** and **g** were performed with biological replicates; **c–f** were performed with technical replicates. All experiments were repeated independently three times. All results are shown as mean ($n = 5$ for **g**; $n = 3$ for all other experiments) \pm s.d.

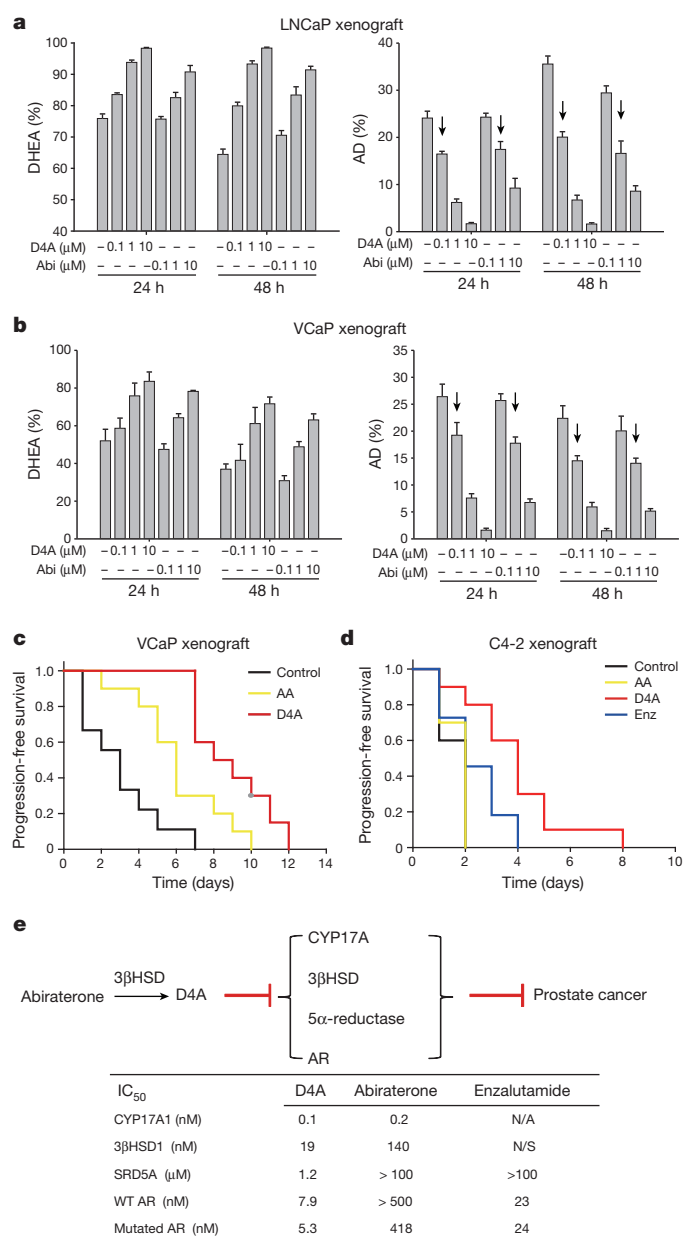


Figure 4 | D4A inhibits xenograft steroidogenesis and growth. **a, b**, D4A inhibits steroidogenesis in LNCaP and VCaP xenografts. Xenografts were grown subcutaneously in NSG mice. Xenografts reaching ~1,000 mm³ were minced and treated with [³H]DHEA, plus Abi or D4A (0.1, 1 and 10 μM). DHEA and AD were separated and quantified by HPLC. Experiments were performed with biological replicates (*n* = 3 xenograft tissues) and results are shown as mean ± s.d. Arrows denote AD percentage for 0.1 μM D4A and 1 μM Abi treatment groups. **c**, D4A is more potent than Abi for blocking VCaP xenograft progression. Time to xenograft progression (>20% increase in tumour volume) is shown for vehicle (control; *n* = 9 mice), Abi acetate (AA; *n* = 10) and D4A (*n* = 10) treatment groups. All three treatment groups differ significantly from one another: control versus AA, *P* = 0.02; control versus D4A, *P* < 0.001; AA versus D4A, *P* = 0.01. **d**, D4A (*n* = 10) is more potent than AA (*n* = 10) and Enz (*n* = 11) for blocking C4-2 xenograft progression: control versus D4A, *P* < 0.001; AA versus D4A, *P* = 0.01; Enz versus D4A, *P* = 0.02. **e**, Schematic of D4A activities and IC₅₀ in the androgen pathway and comparisons with Abi and Enz.

and comparing treatment groups with the log-rank test. Progression was significantly delayed in the D4A group compared with the Abi acetate group (*P* = 0.011). Fold-change in tumour volume is shown for each treatment group in Extended Data Fig. 6. We also compared

xenograft growth using the same method with the C4-2 cell line model. D4A treatment increased progression-free survival compared with Abi acetate and enzalutamide (Fig. 4d). In serum collected from D4A-treated mice at the end of the xenograft study there was no detectable increase in deoxycorticosterone, which is the mineralocorticoid that is most highly elevated in patients treated with Abi acetate, causing hypertension and hyperkalaemia (Extended Data Fig. 7)^{6,16}. Figure 4e depicts the multiple points in the androgen pathway at which conversion of Abi by 3βHSD to D4A in patients impinges on AR signalling and prostate cancer progression, and the relative potencies of D4A, Abi and enzalutamide.

The next-generation hormonal therapies, Abi and enzalutamide, each have a single predominant target (CYP17A1 and AR, respectively). These drugs have clinically validated that androgen synthesis and AR stimulation are both essential components required to spur the development and progression of CRPC. After oral administration, Abi acetate is hydrolysed and thereby converted to Abi, which is thought to be the major active agent by way of blocking CYP17A1. The major recognized metabolites of Abi result from hepatic CYP3A4 and SULT2A1 processing, forming the N-oxide of Abi and Abi sulfate, respectively. Neither of these modifications affects the Δ⁵, 3β-hydroxyl-structure of the steroid backbone. In contrast, conversion to D4A modifies the steroidal structure to one that more robustly engages with AR, SRD5A and 3βHSD, thereby blocking androgen signalling at all these steps, while retaining CYP17A1 inhibition. The clinical significance of conversion of Abi to D4A and effects on individual components of the androgen pathway in patients must be viewed in light of pharmacokinetic studies that show a maximum serum concentration (*C*_{max}) of approximately 1 μM and also wide inter-patient variability¹⁶. Furthermore, our findings suggest that D4A also has much more potent anti-tumour activity against CRPC when directly compared with Abi.

The potential clinical utility of treating patients with CRPC directly with D4A is dependent on the underlying mechanisms of resistance to Abi, which have not been fully elucidated, and the clinical settings in which the benefit from Abi is exhausted. The evidence suggests that sustained AR signalling characterizes at least a subset of Abi resistance cases. For example, increased AR copy number in CRPC is associated with absence of clinical response to Abi¹⁷, and increased AR protein expression appears to occur upon the development of acquired clinical resistance to Abi¹⁸. Although Abi is a potent CYP17A1 inhibitor, studies of urinary steroid metabolites in patients demonstrate that androgen synthesis inhibition is incomplete, raising the possibility of sustained steroidogenesis as a mechanism of resistance^{5,19}. Clinically, the combination of CYP17A1 inhibition and a potent AR antagonist appears to confer more potent androgen signalling inhibition in patients with CRPC²⁰.

Circulating concentrations of D4A in patients treated with Abi acetate appear to be quite variable. In contrast to hepatic Abi metabolites, it is probable that conversion of Abi to D4A in peripheral tissues leads to D4A concentrations that are higher in peripheral tissues (such as CRPC) than are present in serum. The effects of D4A on androgen signalling in CRPC, in particular distal steps in DHT synthesis and activity as an AR antagonist, may therefore be underestimated on the basis of serum concentrations alone. Nonetheless, D4A levels in peripheral tissues and the precise contribution of D4A to the clinical activity of Abi have yet to be determined.

Finally, our results raise the possibility that there may be a previously unappreciated class effect of steroidal versus non-steroidal CYP17A1 inhibitors. In contrast to Abi acetate, TAK-700, a non-steroidal CYP17A1 inhibitor, failed to prolong survival in metastatic CRPC²¹. It is possible that the absence of active downstream steroidal metabolites may have contributed to these findings. This issue should be considered as other steroidal and non-steroidal CYP17A1 inhibitors undergo further clinical investigation.

In conclusion, we have identified a novel Abi metabolite that is present in patients with CRPC treated with Abi acetate and has more potent anti-tumour activity than the parent drug. Conversion to D4A may be responsible for some of the clinical activity observed with the use of Abi. We suggest that treatment with D4A is likely to result in a greater clinical benefit than Abi.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 October 2014; accepted 10 March 2015.

Published online 1 June 2015.

- Titus, M. A., Schell, M. J., Lih, F. B., Tomer, K. B. & Mohler, J. L. Testosterone and dihydrotestosterone tissue levels in recurrent prostate cancer. *Clin. Cancer Res.* **11**, 4653–4657 (2005).
- Montgomery, R. B. *et al.* Maintenance of intratumoral androgens in metastatic prostate cancer: a mechanism for castration-resistant tumor growth. *Cancer Res.* **68**, 4447–4454 (2008).
- Chang, K. H. *et al.* A gain-of-function mutation in DHT synthesis in castration-resistant prostate cancer. *Cell* **154**, 1074–1084 (2013).
- Knudsen, K. E. & Penning, T. M. Partners in crime: deregulation of AR activity and androgen synthesis in prostate cancer. *Trends Endocrinol. Metab.* **21**, 315–324 (2010).
- Chang, K. H. & Sharifi, N. Prostate cancer—from steroid transformations to clinical translation. *Nature Rev. Urol.* **9**, 721–724 (2012).
- de Bono, J. S. *et al.* Abiraterone and increased survival in metastatic prostate cancer. *N. Engl. J. Med.* **364**, 1995–2005 (2011).
- Ryan, C. J. *et al.* Abiraterone in metastatic prostate cancer without previous chemotherapy. *N. Engl. J. Med.* **368**, 138–148 (2013).
- Scher, H. I. *et al.* Increased survival with enzalutamide in prostate cancer after chemotherapy. *N. Engl. J. Med.* **367**, 1187–1197 (2012).
- Beer, T. M. *et al.* Enzalutamide in metastatic prostate cancer before chemotherapy. *N. Engl. J. Med.* **371**, 424–433 (2014).
- Li, R. *et al.* Abiraterone inhibits 3 β -hydroxysteroid dehydrogenase: a rationale for increasing drug exposure in castration-resistant prostate cancer. *Clin. Cancer Res.* **18**, 3571–3579 (2012).
- Ferraldeschi, R., Sharifi, N., Auchus, R. J. & Attard, G. Molecular pathways: inhibiting steroid biosynthesis in prostate cancer. *Clin. Cancer Res.* **19**, 3353–3359 (2013).
- Garrido, M. *et al.* A-ring modified steroidal azoles retaining similar potent and slowly reversible CYP17A1 inhibition as abiraterone. *J. Steroid Biochem. Mol. Biol.* **143**, 1–10 (2014).
- Chang, K. H. *et al.* Dihydrotestosterone synthesis bypasses testosterone to drive castration-resistant prostate cancer. *Proc. Natl Acad. Sci. USA* **108**, 13728–13733 (2011).
- Richards, J. *et al.* Interactions of abiraterone, eplerenone, and prednisolone with wild-type and mutant androgen receptor: a rationale for increasing abiraterone exposure or combining with MDV3100. *Cancer Res.* **72**, 2176–2182 (2012).
- Hodgson, M. C. *et al.* The androgen receptor recruits nuclear receptor CoRepressor (N-CoR) in the presence of mifepristone via its N and C termini revealing a novel molecular mechanism for androgen receptor antagonists. *J. Biol. Chem.* **280**, 6511–6519 (2005).
- Attard, G. *et al.* Phase I clinical trial of a selective inhibitor of CYP17, abiraterone acetate, confirms that castration-resistant prostate cancer commonly remains hormone driven. *J. Clin. Oncol.* **26**, 4563–4571 (2008).
- Carreira, S. *et al.* Tumor clone dynamics in lethal prostate cancer. *Sci. Transl. Med.* **6**, 254ra125 (2014).
- Balk, S. Emerging developments in abiraterone resistance. In *Proc. American Association for Cancer Research Prostate Cancer Foundation Conference on Advances in Prostate Cancer Research*. Abstr. IA11 (American Association for Cancer Research, 2014).
- Attard, G. *et al.* Clinical and biochemical consequences of CYP17A1 inhibition with abiraterone given with and without exogenous glucocorticoids in castrate men with advanced prostate cancer. *J. Clin. Endocrinol. Metab.* **97**, 507–516 (2012).
- Efstathiou, E. *et al.* Enzalutamide (ENZA) in combination with abiraterone acetate (AA) in bone metastatic castration resistant prostate cancer (mCRPC). *J. Clin. Oncol.* **32**, abstr. 5000 (2014).
- Fizazi, K. *et al.* Phase III, randomized, double-blind, multicenter trial comparing orteronel (TAK-700) plus prednisone with placebo plus prednisone in patients with metastatic castration-resistant prostate cancer that has progressed during or following docetaxel-based therapy: ELM-PC 5. *J. Clin. Oncol.* **33**, 723–731 (2015).
- Papari-Zareei, M., Brandmaier, A. & Auchus, R. J. Arginine 276 controls the directional preference of AKR1C9 (rat liver 3 α -hydroxysteroid dehydrogenase) in human embryonic kidney 293 cells. *Endocrinology* **147**, 1591–1597 (2006).
- Li, Z., Nie, F., Wang, S. & Li, L. Histone H4 Lys 20 monomethylation by histone methylase SET8 mediates Wnt target gene activation. *Proc. Natl Acad. Sci. USA* **108**, 3116–3123 (2011).
- Evaul, K., Li, R., Papari-Zareei, M., Auchus, R. J. & Sharifi, N. 3 β -hydroxysteroid dehydrogenase is a possible pharmacological target in the treatment of castration-resistant prostate cancer. *Endocrinology* **151**, 3514–3520 (2010).
- Tran, C. *et al.* Development of a second-generation antiandrogen for treatment of advanced prostate cancer. *Science* **324**, 787–790 (2009).
- Kushnir, M. M. *et al.* Liquid chromatography-tandem mass spectrometry assay for androstenedione, dehydroepiandrosterone, and testosterone with pediatric and adult reference intervals. *Clin. Chem.* **56**, 1138–1147 (2010).

Acknowledgements This work has been supported in part by funding from a Howard Hughes Physician-Scientist Early Career Award (to N.S.), by the Prostate Cancer Foundation (to N.S.), by an American Cancer Society Research Scholar Award (12-038-01-CCE; to N.S.), grants from the US Army Medical Research and Materiel Command (PC080193 to N.S. and PC121382 to Z.L.), and additional grants from the National Cancer Institute (R01CA168899, R01CA172382, and R01CA190289; to N.S.).

Author Contributions Z.L. performed gene expression, metabolism, chromatin immunoprecipitation (ChIP) and mouse xenograft studies. A.B., M.A. and D.B. performed mass spectrometry studies. J.A.G. and R.D. participated in clinical studies. J.L. and S.K.U. performed enzymology studies, and S.K.U. also performed chemical syntheses. Z.L., R.J.A. and N.S. designed the studies and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.S. (sharifn@ccf.org).

METHODS

No statistical methods were used to predetermine sample size.

Cell lines. LNCaP and VCaP cells were purchased from the American Type Culture Collection and maintained in RPMI-1640 with 10% FBS. LAPC4 cells were provided by C. Sawyers and grown in Iscove's Modified Dulbecco's Medium with 10% FBS. C4-2 cells were provided by L. Chung and maintained in RPMI-1640 with 10% FBS. All experiments using LNCaP and VCaP were performed in plates coated with poly-DL-ornithine (Sigma-Aldrich). A 293-cell line stably expressing human CYP17A1 was generated by transfection with (pcDNA3)-CYP17 (a gift from W. Miller) and selection with G418 as described²². Cell lines were authenticated by DDC Medical and determined to be negative for mycoplasma with primers 5'-ACACCATGGGAGCTGGTAAT-3' and 5'-GTTTCATCGACTTTCAGACCCAAGGCAT-3'.

Chemicals. Abi acetate was purchased from Medkoo Biosciences. Abi and D4A were synthesized as described previously¹⁰. Enzalutamide was obtained from Medivation.

Steroid metabolism. Cell line metabolism. Cells were seeded and incubated in 12-well plates with 0.2 million cells per well for ~24 h and then incubated with a mixture of radioactive (³H-labelled) and non-radioactive androgens (final concentration, 100 nM; ~1,000,000 counts per minute (c.p.m.) per well; PerkinElmer) at 37 °C. Aliquots of medium were collected at the indicated times. Collected medium was treated with 1,000 units of β -glucuronidase (*Helix pomatia*; Sigma-Aldrich) at 65 °C for 2 h, extracted with 860 μ l ethyl acetate:isooctane (1:1), and concentrated under nitrogen gas.

Xenograft metabolism. Ten million LNCaP or VCaP cells were injected subcutaneously with Matrigel into surgically orchiectomized NSG mice that were implanted with 5 mg 90-day sustained-release DHEA pellets (Innovative Research of America). Xenografts were harvested when they reached 1,000 mm³, minced, and cultured in DMEM with 10% FBS at 37 °C with a mixture of radioactive (³H-labelled) and non-radioactive androgens, or Abi, when they reached 1,000 mm³. Each xenograft was analysed in triplicate in at least three independent experiments. Aliquots of medium were collected at the indicated times. Collected medium was treated with 1,000 units of β -glucuronidase at 65 °C for 2 h, extracted with 860 μ l ethyl acetate:isooctane (1:1) and concentrated under nitrogen gas.

HPLC analysis was performed on a Waters 717 Plus HPLC or an Agilent 1260 HPLC. Dried samples were reconstituted in 50% methanol and injected into the HPLC. Steroids and drug metabolites were separated on a Kinetex 100 mm \times 2.1 mm, 2.6 μ m particle size C₁₈ reverse-phase column (Phenomenex) using a methanol/water gradient at 30 °C. The column effluent was analysed using a dual-wavelength ultraviolet-visible detector set at 254 nm or β -RAM model 3 in-line radioactivity detector (IN/US Systems) using Liquescent scintillation cocktail (National Diagnostics). Alternatively, dried samples were applied to plastic-backed silica gel plates and separated by TLC using a mobile phase of 3:1 chloroform:ethyl acetate, followed by exposure of the plates to a phosphorimager screen and quantitated with a Storm model 860 phosphorimager (Applied Biosystems). All HPLC and TLC studies were conducted in triplicate and repeated at least three times in independent experiments. Results are shown as mean \pm s.d.

Gene expression. Cells were starved with phenol-red-free and serum-free medium for at least 48 h and treated with the indicated drugs and/or androgens. RNA was extracted with a GenElute Mammalian Total RNA miniprep kit (Sigma-Aldrich). cDNA was synthesized from 1 μ g RNA in a reverse transcription reaction using the iScript cDNA Synthesis Kit (Bio-Rad). qPCR analysis was conducted in triplicate with primers for PSA, *TMPRSS2*, *FKBP5* and *RPLPO* described previously³, with an ABI 7500 Real-Time PCR machine (Applied Biosystems) using iTaq Fast SYBR Green Supermix with ROX (Bio-Rad) in 96-well plates at a final reaction volume of 20 μ l. Accurate quantitation of each messenger RNA was achieved by normalizing the sample values to *RPLPO* and to vehicle-treated cells.

ChIP assay. LNCaP cells were serum-starved for at least 48 h and treated with indicated drugs and DHT for 3 h. A ChIP assay was performed with an anti-AR antibody (Santa Cruz, sc-816) as described previously²³. All of the precipitated DNA samples were quantified by qPCR and normalized to input DNA. All ChIP experiments were performed at least three times. Primers used for ChIP experiments were from experiments previously described²⁴. Primer sequences are as follows: PSA: 5'-TGGGACAACCTTGCAACCTG-3' and 5'-CCAGAGTAGGTCGTGTTTCAATCCA-3'; *FKBP5*: 5'-CCCCCTATTTAATCGGAGTAC-3' and 5'-TTTTGAAGAGCACAGAACACCT-3'; *TMPRSS2*: 5'-TGGTCCTGGATGATAAAAAAGTTT-3' and 5'-GACATACGCCCCACAACAGA-3'.

Mouse xenograft studies. Male NSG mice, 6–8 weeks of age, were obtained from the Cleveland Clinic Biological Resources Unit facility. All mouse studies were conducted under a protocol approved by the Cleveland Clinic Institutional

Animal Care and Use Committee. Sample size was determined on the basis of our prior studies of steroidogenesis inhibition in xenograft models of CRPC^{3,13}. Criteria for progression were determined on the basis of similar criteria used for clinical progression. Mice were surgically orchiectomized and implanted with a 5 mg 90-day sustained-release DHEA pellet (Innovative Research of America) to mimic CRPC in the context of human adrenal physiology. Two days later, 10⁷ VCaP or C4-2 cells were injected subcutaneously with matrigel. Once tumours reached 300 mm³ (length \times width \times height \times 0.52), mice were arbitrarily (but not strictly randomized) assigned to vehicle (n = 9 or 10 mice for VCaP and C4-2 respectively), Abi acetate (n = 10 mice for both cell lines), D4A (n = 10 mice for both cell lines) or enzalutamide (n = 11 for C4-2) treatment groups. Abi acetate and D4A (0.5 mmol per kg per day in 0.1 ml 5% benzyl alcohol and 95% safflower oil solution) were administered via 5 ml per kg intraperitoneal injection every day for up to 15 days. Control groups were administered 0.1 ml 5% benzyl alcohol and 95% safflower oil solution via intraperitoneal injection every day. Mice in the enzalutamide group were fed with enzalutamide in chow every day (10 mg per kg per day)²⁵. Treatment was not blinded to the investigator. Tumour volume was measured daily, and time to increase in tumour volume by 20% was determined. Mice were killed at treatment day 15 or when the tumour size was twofold greater than baseline. The significance of the difference between treatment groups was assessed by Kaplan–Meier survival analysis using a log-rank test in SigmaStat 3.5. A two-tailed Student's *t*-test was used to determine significance in Extended Data Fig. 6.

Enzyme assays. To test D4A as an inhibitor of 3 β HSD, enzyme assays were performed as described previously¹⁰. Briefly, incubations were prepared with recombinant human 3 β HSD1 or 3 β HSD2 (in yeast microsomes, 45 or 2.5 μ g protein per incubation, respectively), [³H]pregnenolone (100,000 c.p.m., 1–20 μ M), and D4A (5–20 μ M) or ethanol vehicle in 0.5 ml of potassium phosphate buffer (pH 7.4). After a pre-incubation at 37 °C for 1–3 min, NAD⁺ (1 mM) was added, and the incubation was conducted at 37 °C for 20 min. The reaction was stopped by addition of 1 ml ethyl acetate:isooctane (1:1), and the steroids were then extracted into the organic phase and dried. The steroids in the dried extracts were resolved by HPLC and quantified by in-line scintillation counting.

AR competition assay. Cells were cultured in serum-free medium for 48 h and then treated with [³H]R1881 and the indicated concentrations of D4A, Abi, enzalutamide, R1881 or bicalutamide for 30 min. Cells were washed with 1 \times PBS four times and 0.9% NaCl solution twice before lysis with RIPA buffer. Intracellular radioactivity was measured with a Beckman Coulter LS6000IC liquid scintillation counter and normalized to the protein concentration as detected with a Wallac Victor2 1420 Multilabel counter (PerkinElmer).

Cell proliferation assay. LNCaP cells (0.1 million cells per millilitre) were seeded in 96-well plates and cultured in phenol-red-free RPMI-1640 mediums plus 10% charcoal stripped FBS with androgens and/or drugs. Medium was changed every other day. After 2, 4 or 6 days, cells were collected and lysed after treatment. Growth was then determined on the basis of DNA content as detected by Hoechst stain and a Wallac Victor2 1420 Multilabel counter (PerkinElmer). A two-tailed Student's *t*-test was used to determine significance.

Mass spectrometry. Patient serum collection and drug extraction. Twelve patients with CRPC undergoing treatment with Abi acetate were consented under an Institutional Review Board-approved protocol (case 7813). Blood was collected using Vacutainer Plus serum blood collection tubes (BD367814, Becton Dickinson). Blood was collected between 2 and 14 h after the 1,000 mg daily dose of Abi acetate was administered. Blood was allowed to clot and tubes were centrifuged at 2,500 r.p.m. for 10 min. Serum aliquots were frozen at –80 °C until processing. Drug metabolites and internal standard (d₄-cortisol, CDN Isotopes) were extracted from 200 μ l of patient serum with methyl tert-butyl ether (Sigma Aldrich), evaporated under a stream of nitrogen gas and reconstituted in methanol before mass spectrometry analysis.

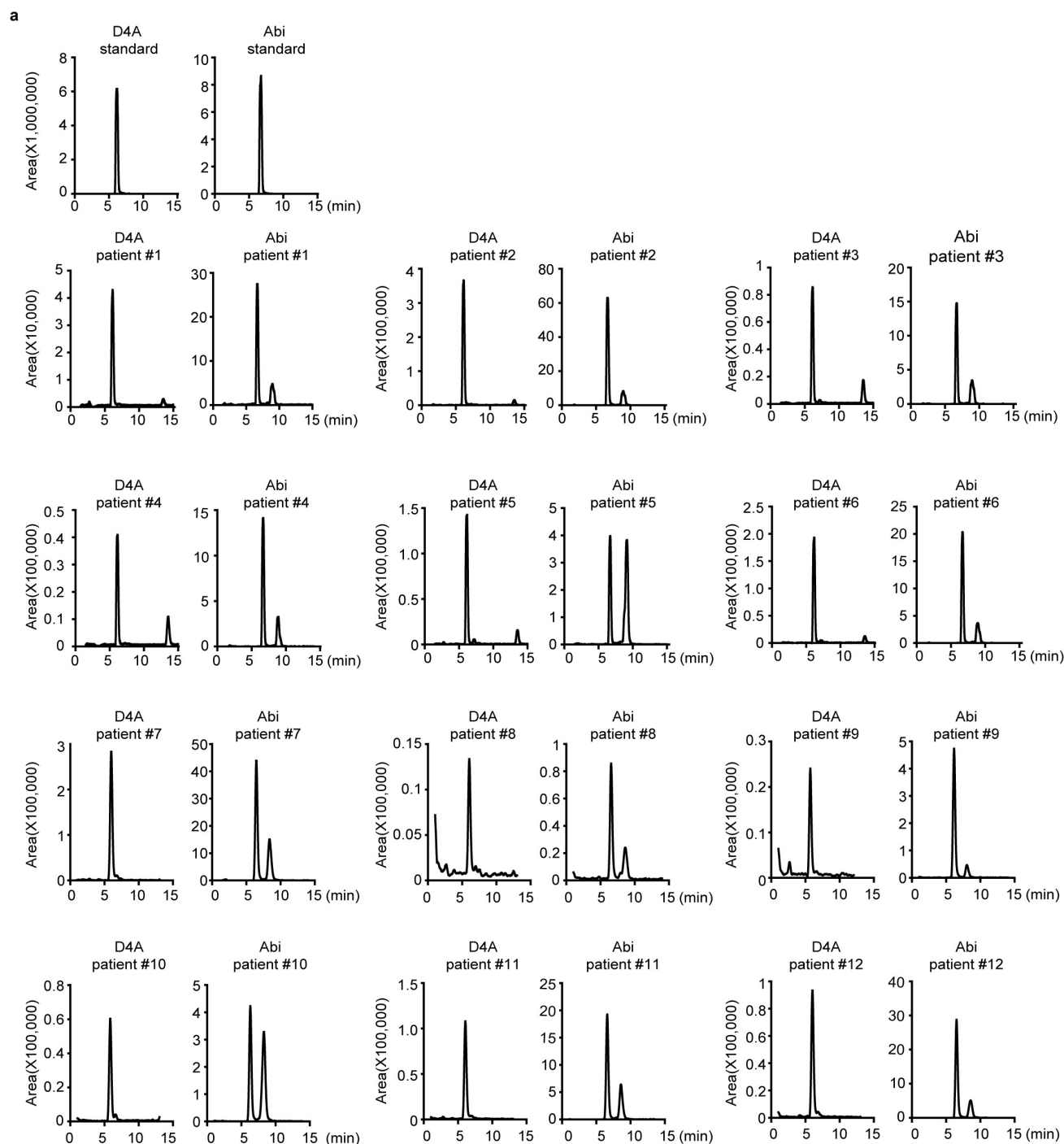
Mouse serum collection, derivatization and extraction. At the completion of the mouse xenograft study, mouse serum was collected for steroid analysis. Twenty microlitres of serum and internal standard (d₃-deoxycorticosterone) was derivatized with hydroxylamine (Sigma Aldrich)²⁶. Steroids were extracted with methyl tert-butyl ether (Sigma Aldrich), evaporated under a stream of nitrogen gas and reconstituted in methanol:water (50:50) before mass spectrometry analysis.

Stable isotope dilution liquid chromatography–mass spectrometry analysis. Mouse serum analysis. Samples were analysed on a Thermo TSQ Quantiva-Prelude SPLC system (Thermo Scientific) with Aria MX 2.1 and Tracefinder 3.2.368.22 software for instrument controls and quantitation. Analyte separation was achieved with an Accucore 50 mm \times 3 mm, 2.6 μ m C18 column (Thermo Scientific) using a mobile phase of liquid chromatography–mass spectrometry-grade methanol and water (Thermo Scientific), a gradient of 25–100% methanol and a flow rate of 0.6 ml min^{–1}. Steroids were ionized by electrospray ionization and in positive ion mode. Multiple reaction monitoring was used to follow mass

transitions for the oximes of deoxycorticosterone (m/z : 361.2/124.1) and d_8 -deoxycorticosterone (m/z : 369.4/128.1) (Steraloids). Concentrations were determined using stable isotope dilution analysis.

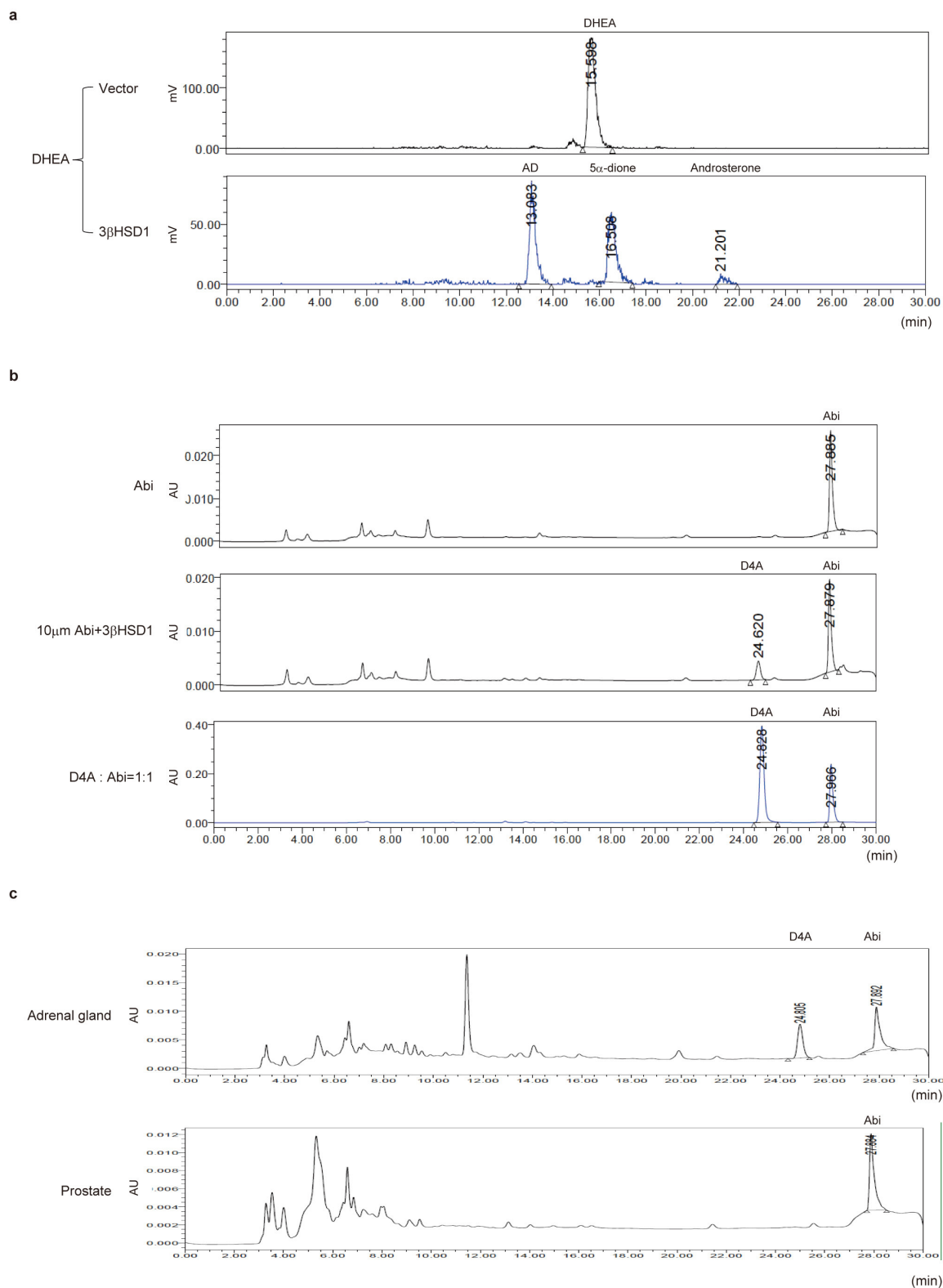
Patient serum analysis. Blood was obtained from prostate cancer patients with consent under a protocol approved by the Cleveland Clinic Institutional Review Board. Samples were analysed on an ultra-HPLC station (Shimadzu) with a DGU-20A3R degasser, two LC-30AD pumps, a SIL-30AC autosampler, a CTO-10A column oven and a CBM-20A system controller in tandem with a QTRAP 5500 mass spectrometer (AB Sciex). The mobile phase consisted of liquid chromato-

graphy-mass spectrometry grade (Fisher) methanol:acetonitrile:water (44:36:20). Separation of drug metabolites was achieved using a Zorbax Eclipse plus 150 mm \times 2.1 mm, 3.5 μ m C18 column (Agilent) at a flow rate of 0.2 ml min⁻¹. Drug metabolites were ionized using electrospray ionization in positive ion mode. Multiple reaction monitoring was used to follow mass transitions for D4A (m/z : 348.2/156.3), abiraterone (m/z : 350.3/156.1) and d_4 -cortisol (m/z : 367.1/121.1). Standard curves were generated using human serum spiked with known concentrations of each metabolite to enable determination of unknown concentrations in patient samples.



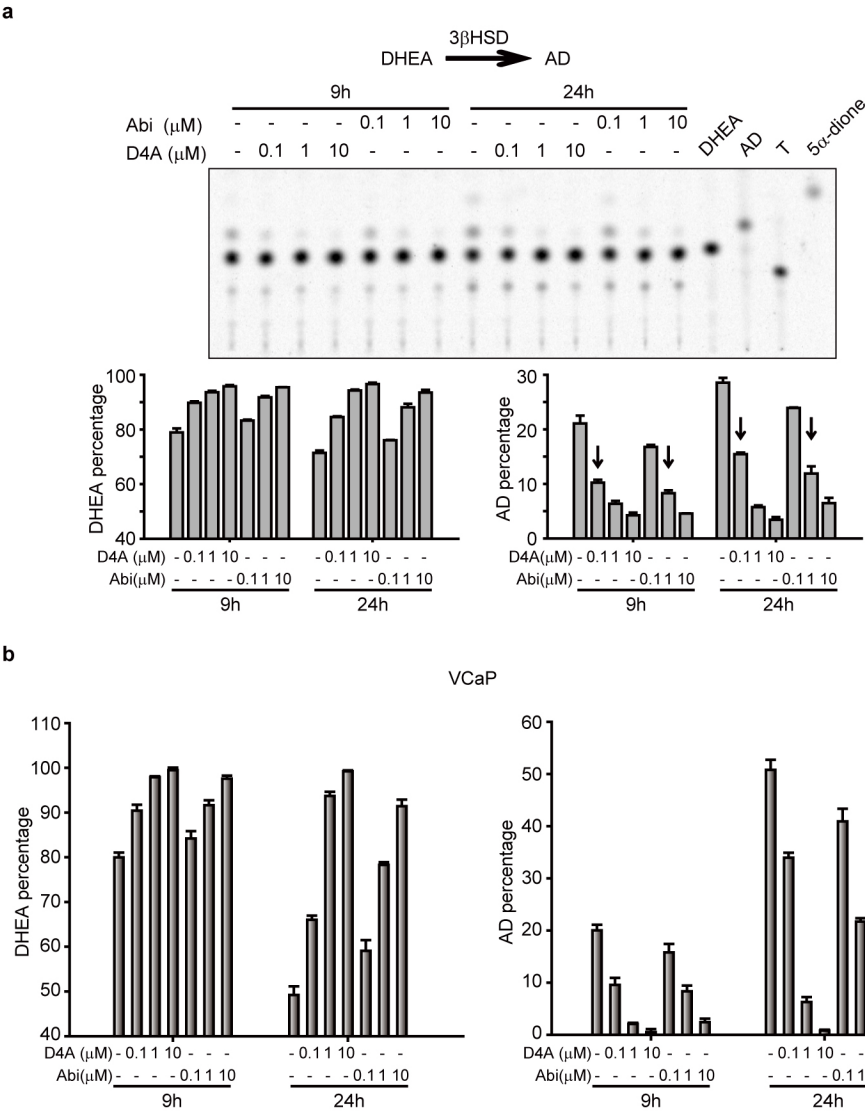
Extended Data Figure 1 | D4A is detectable in patients with prostate cancer treated with Abi acetate. a, Mass spectrometry tracings of Abi and D4A in the serum of 12 patients treated with Abi acetate. Blood was drawn between 2 and

14 h after the administration of the 1,000 mg daily dose. **b,** Duration of Abi acetate therapy and time from last dose to blood draw for individual patients.



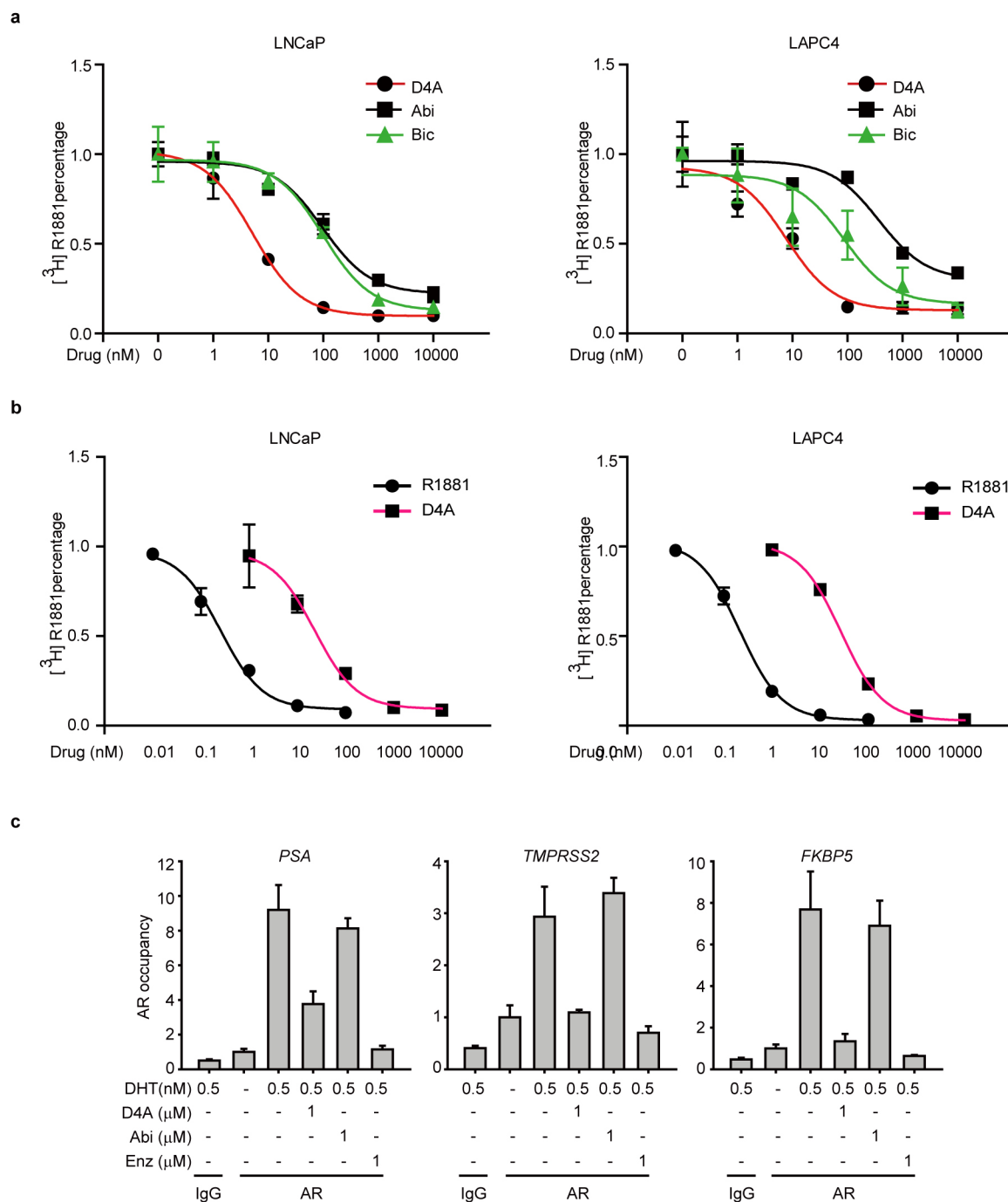
Extended Data Figure 2 | 3βHSD converts Abi to D4A. **a**, 3βHSD1 expression permits catalysis of DHEA to AD. LAPC4 cells were transfected with 3βHSD1 or vector and treated with [3 H]DHEA. Medium was collected 24 h later and androgens were separated and quantified by HPLC. **b**, 3βHSD1 expression allows conversion of Abi to D4A. LAPC4 cells were transfected with 3βHSD1 or vector and then treated with Abi. Medium was collected after 24 h

and D4A and Abi were separated by HPLC. **c**, 3βHSD enzymatic activity present in the mouse adrenal gland but not prostate gland converts Abi to D4A. Mouse adrenal and prostate glands were harvested and minced before culturing in the presence of media containing Abi. Medium was collected after 24 h for separation and quantitation of D4A and Abi by HPLC.



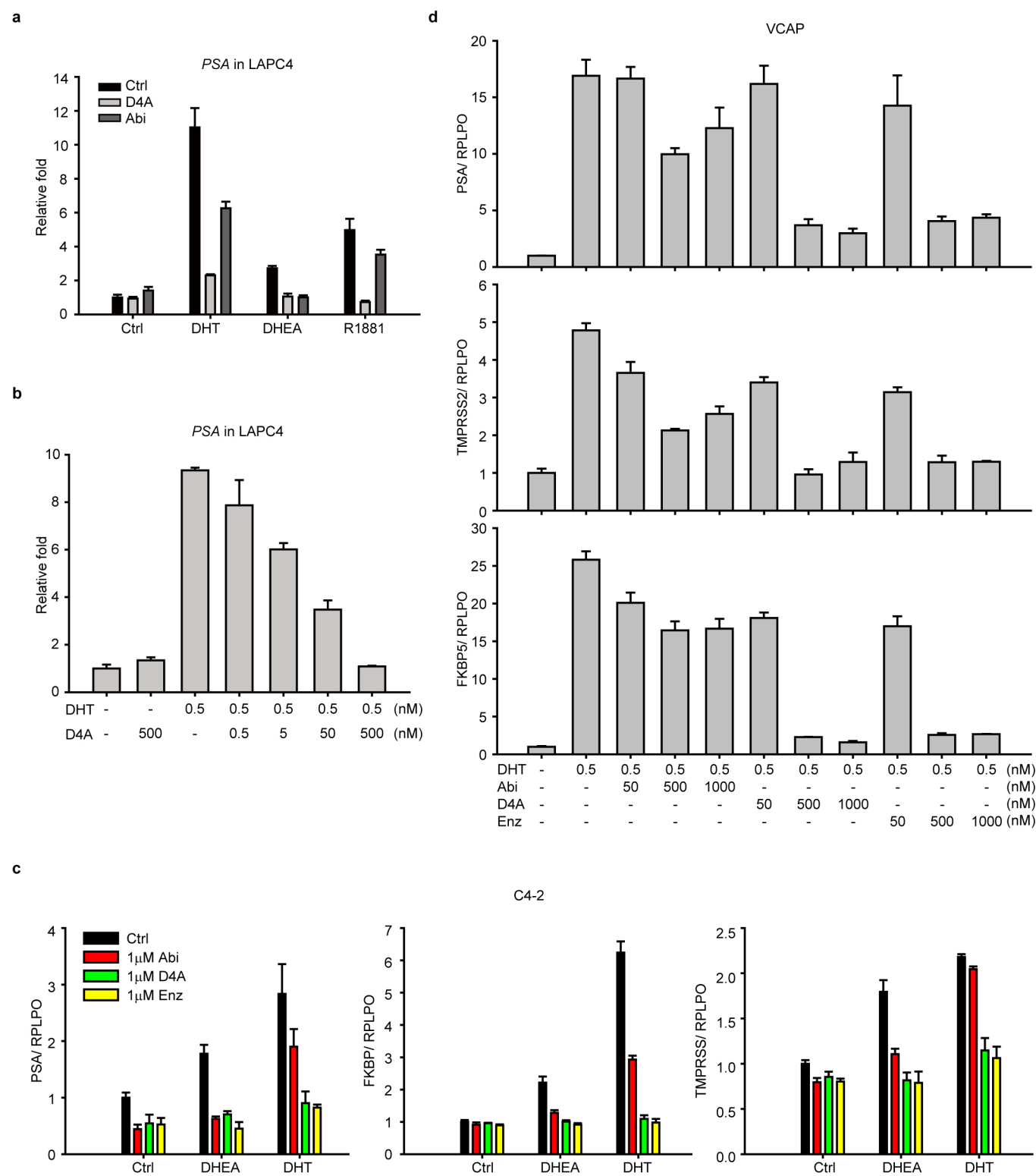
Extended Data Figure 3 | D4A inhibits $3\beta\text{HSD1}$ activity. **a**, D4A inhibits $3\beta\text{HSD1}$ activity in LNCaP. Cells were treated with [^3H]DHEA (DHEA: 100nM; [^3H]DHEA, 1,000,000 c.p.m. per well) with 0.1, 1 or 10 μM D4A and Abi, for 9 and 24 h. DHEA and AD were separated and quantified by TLC. ImageJ was used to quantify steroids. For ease of comparison, arrows denote

AD percentage for 0.1 μM D4A and 1 μM Abi treatment groups. **b**, VCaP cells were treated with [^3H]DHEA and the indicated concentrations of D4A and Abi. The percentages of DHEA and AD were determined by HPLC. Experiments were performed with biological replicates ($n = 3$) and results are shown as mean \pm s.d.



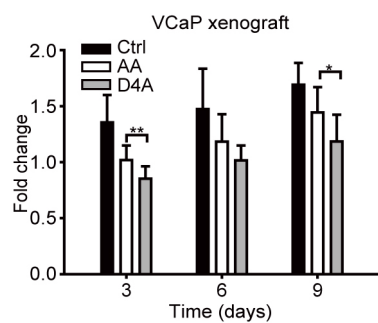
Extended Data Figure 4 | D4A has a higher affinity for both mutant-type AR (LNCaP cells) and wild-type AR (LAPC4 cells) than abiraterone (Abi) and bicalutamide (Bic) and inhibits AR chromatin occupancy better than Abi. **a**, Competition plots for D4A, Abi and Bic. **b**, Competition plots for unlabelled R1881 and D4A. Displacement of [^3H]R1881 is described in Methods. Experiments were performed with biological replicates ($n = 3$) and results are shown as mean \pm s.d. **c**, D4A inhibition of AR chromatin

occupancy is superior to Abi. LNCaP cells were treated with the indicated concentrations of DHT, D4A, Abi and enzalutamide (Enz) for 3 h. AR chromatin occupancy for PSA, TMPSR2 and FKBP5 was detected with ChIP. Experiments were performed with technical replicates ($n = 3$) and results are shown as mean \pm s.d. All experiments were repeated independently at least three times.

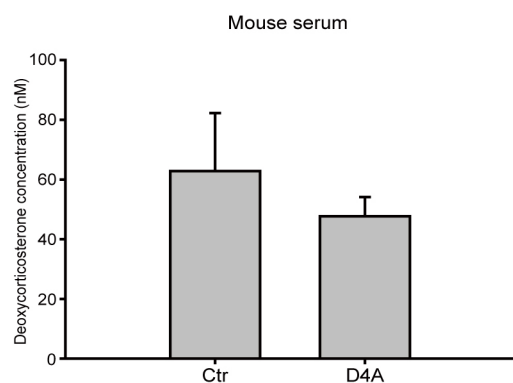


Extended Data Figure 5 | D4A inhibits expression of androgen-responsive genes. **a**, D4A inhibits PSA expression in LAPC4 cells. Cells were treated with DHT (0.5 nM), DHEA (40 nM) or R1881 (0.1 nM) with or without Abi or D4A (1 μ M) for 24 h. **b**, D4A inhibits PSA expression in LAPC4 in a dose-dependent manner. **c**, D4A inhibits AR target gene expression in C4-2 cells. Cells were treated with vehicle control (Ctrl), DHT (0.5 nM) or DHEA (40 nM)

with or without Abi (1 μ M), D4A (1 μ M) or Enz (1 μ M) for 24 h. **d**, D4A is comparable to Enz in inhibiting DHT-induced target gene expression in VCAp cells. Gene expression was assessed in triplicate, detected by qPCR and normalized to *RPLPO*. Experiments were performed with technical replicates ($n = 3$) and results are shown as mean \pm s.d. All experiments were repeated independently at least three times.



Extended Data Figure 6 | D4A impedes VCaP xenograft growth. $*P < 0.05$ and $**P < 0.01$ for the difference between D4A ($n = 10$ mice) and AA ($n = 10$ mice) treatment groups. $N = 9$ mice for the control group.



Extended Data Figure 7 | D4A does not increase deoxycorticosterone concentrations. Serum of mice undergoing long-term treatment with D4A was collected and deoxycorticosterone concentrations were determined by liquid chromatography–mass spectrometry. Compared with control mice injected with vehicle, D4A does not increase deoxycorticosterone concentrations. For both D4A and control groups, $n = 9$ biological replicates (mice).

Improving survival by exploiting tumour dependence on stabilized mutant p53 for treatment

E. M. Alexandrova^{1*}, A. R. Yallowitz^{1*}, D. Li¹, S. Xu¹, R. Schulz², D. A. Proia³, G. Lozano⁴, M. Dobbelstein² & U. M. Moll^{1,2}

Missense mutations in p53 generate aberrant proteins with abrogated tumour suppressor functions that can also acquire oncogenic gain-of-function activities that promote malignant progression, invasion, metastasis and chemoresistance^{1–5}. Mutant p53 (mutp53) proteins undergo massive constitutive stabilization specifically in tumours, which is the key requisite for the acquisition of gain-of-functions activities^{6–8}. Although currently 11 million patients worldwide live with tumours expressing highly stabilized mutp53, it is unknown whether mutp53 is a therapeutic target *in vivo*. Here we use a novel mutp53 mouse model expressing an inactivatable R248Q hotspot mutation (floxQ) to show that tumours depend on sustained mutp53 expression. Upon tamoxifen-induced mutp53 ablation, allotransplanted and autochthonous tumours curb their growth, thus extending animal survival by 37%, and advanced tumours undergo apoptosis and tumour regression or stagnation. The HSP90/HDAC6 chaperone machinery, which is significantly upregulated in cancer compared with normal tissues, is a major determinant of mutp53 stabilization^{9–12}. We show that long-term HSP90 inhibition significantly extends the survival of mutp53 Q/– (R248Q allele²) and H/H (R172H allele³) mice by 59% and 48%, respectively, but not their corresponding p53^{–/–} (also known as *Trp53*^{–/–}) littermates. This mutp53-dependent drug effect occurs in H/H mice treated with 17DMAG+SAHA and in H/H and Q/– mice treated with the potent Hsp90 inhibitor ganetespib. Notably, drug activity correlates with induction of mutp53 degradation, tumour apoptosis and prevention of T-cell lymphomagenesis. These proof-of-principle data identify mutp53 as an actionable cancer-specific drug target.

Tumour-promoting gain-of-function activities of missense mutp53 alleles are demonstrated by knock-in mouse models^{1–5} and in patients with Li-Fraumeni syndrome carrying germline TP53 missense mutations. Li-Fraumeni syndrome patients with TP53 missense mutations have earlier cancer onset than patients with loss of p53 expression^{13,14}. Strikingly, p53R248Q/+ Li-Fraumeni syndrome patients have higher tumour numbers and shorter tumour-free survival by 10.5 years than p53null/+ patients². A key feature and prerequisite for gain-of-function is that mutp53 proteins undergo massive stabilization specifically in tumours^{6–8}.

RNA-interference-mediated depletion of mutp53 produces marked cytotoxic effects in human cancer cells in culture and xenograft models¹⁵. Notably, mutp53 knockdown in orthotopic allografts of *Kras*^{G12D} pancreatic murine cancer cells dramatically reduced their metastatic ability¹⁶. Together, this provides a strong rationale for testing mutp53-destabilizing cancer therapy *in vivo*. If confirmed, it might establish mutp53 as an important direct drug target in many cancer patients. To validate this rigorously, we generated conditional inactivatable p53 R248Q (floxQ) mice (Extended Data Fig. 1a–c) with a Rosa26CreERT2 (ERT2) knock-in allele, which mirror the constitutive R248Q (Q) mice² in tumour spectrum and survival (Extended Data Fig. 1d). Tamoxifen/4-hydroxytamoxifen (4OHT) treatment

activates CreERT2, causing deletion of the floxQ allele (with up to 90% efficiency) and cell death in primary T-cell lymphoma cultures, but not in various controls (Fig. 1a and Extended Data Fig. 1e). Likewise, transplantation assays into immunocompromised hosts (subcutaneous and tail vein allografts, prophylactic and therapeutic treatments) showed that floxQ deletion markedly inhibited tumour growth *in vivo* (Fig. 1a–d and Extended Data Fig. 1f–h) and prolonged survival of recipients compared with controls (Fig. 1b and Extended Data Fig. 1f). These data strongly suggest tumour dependence on sustained high levels of mutp53.

Importantly, clinically advanced autochthonous tumours in floxQ/– mice responded to mutp53 ablation with regression or stagnation (Fig. 2a–c and Extended Data Fig. 2a). Mechanistically, this was due to marked tumour apoptosis (Fig. 2d), but not cell cycle arrest (Extended Data Fig. 2b). Notably, mutp53 ablation was also associated with strong suppression of lung metastasis, contrasting with large metastatic nodules in controls (Fig. 2e). Moreover, mutp53 ablation in floxQ/– mice with early disease (10 weeks old) (Fig. 2f) extended median overall and T-lymphoma-specific survival by 37% from 128 to 175 days (Fig. 2g and Extended Data Fig. 2c). Notably, the improved overall survival of floxQ/– mice, which normally have a significantly shorter lifespan than p53-null littermates² (Extended Data Fig. 1d), now resembled that of p53-null mice (Extended Data Fig. 2d), while their T-lymphoma-specific survival now extended beyond that of p53-null mice (Extended Data Fig. 2e). This further indicates that tumours driven by mutp53 depend on stabilized mutp53. In support, at endpoint (death), most tumours of all types (17/23, 74%) from floxQ/– mice that were tamoxifen-treated at 10 weeks were again composed of 100% mutp53-overexpressing cells (Fig. 2h and Extended Data Fig. 2f). This indicates strong selective pressure for the small minority of non-recombined mutp53-positive cells outcompeting the majority of recombined cells. It is tempting to speculate that complete allele removal would have further improved survival. Thus, these data establish for the first time that continued expression of stabilized mutp53 is essential for tumour maintenance *in vivo*.

The HSP90 chaperone machinery is highly activated in cancers compared with normal tissues and renders them resistant to proteotoxic stress by supporting proper folding of conformationally aberrant oncoproteins including mutp53 (refs 17, 18). Thus, cancer cells have a far smaller tolerance for HSP90 inhibition. We and others previously showed that HSP90 and its obligatory positive regulator, cytosolic HDAC6, are major determinants of mutp53 stabilization^{9–12}. Importantly, deletion of HSF1, the master transcriptional activator of the inducible heat-shock response including HSP90, dramatically suppresses oncogenicity in mutp53 H/+ mice, but has no effect in p53-null mice^{19,20}. These data clearly indicate that tumorigenicity of the H allele—but not of p53-null—strongly depends on Hsf1-mediated chaperone support, mainly HSP90. 17AAG and its hydrophilic derivative 17DMAG are ansamycin-derived highly specific first-generation

¹Department of Pathology, Stony Brook University, Stony Brook, New York 11794, USA. ²Institute of Molecular Oncology, University of Göttingen, 37077 Göttingen, Germany. ³Synta Pharmaceuticals Corp., Lexington, Massachusetts 02421, USA. ⁴Department of Cancer Genetics, University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA.

*These authors contributed equally to this work.

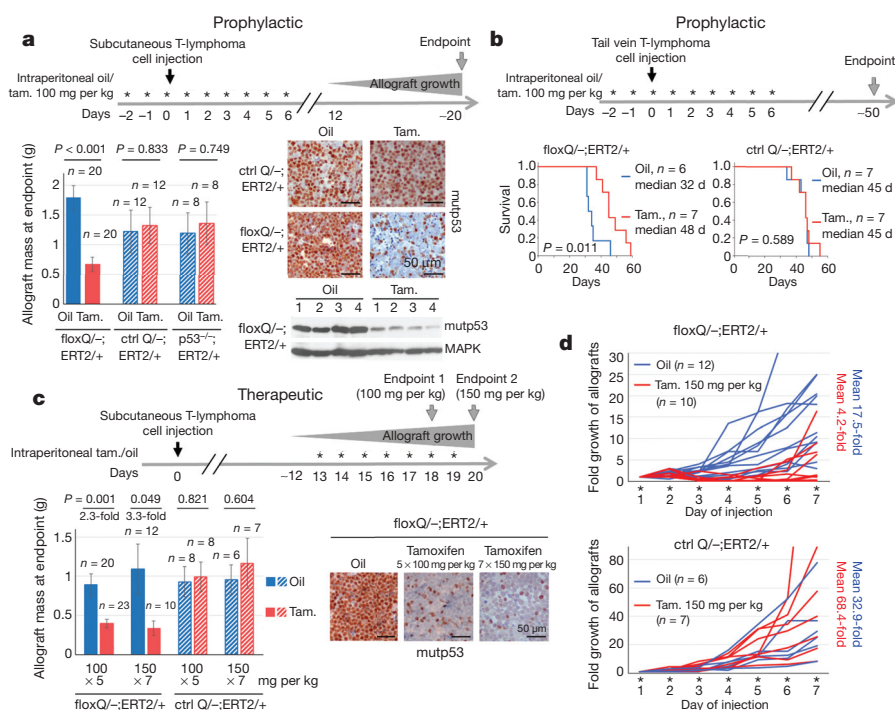


Figure 1 | Genetic ablation of mup53 curbs tumour growth in allografts. **a–d**, Various prophylactic (**a**, **b**) and therapeutic (**c**, **d**) protocols of primary floxQ/- versus Q/- and p53-null T-lymphomas allotransplanted (black arrows on time axes) via subcutaneous (**a**, **c**, **d**) or tail vein (**b**) injections into nude mice, treated with daily intraperitoneal injections of tamoxifen (tam.) or corn oil (stars on time axes). **a**, Experimental diagram, allograft mass, representative tissue immunostaining and immunoblot at endpoint. Unpaired two-tailed Student's *t*-test; mean \pm s.e.m.; *n*, number of allografts. **b**, Deletion of mup53 improves survival of host mice. Kaplan–Meier analysis, log rank test; *n*, number of mice. **c**, Two different therapeutic protocols (100 mg per kg tamoxifen for 5 days versus 150 mg per kg for 7 days) show the dose-dependence of allograft growth on mup53 depletion. Unpaired two-tailed Student's *t*-test; mean \pm s.e.m.; *n*, number of allografts. **d**, Allograft growth using therapeutic protocol (Fig. 1c, endpoint 2) and mean fold changes at endpoint; *n*, number of allografts.

Hsp90 inhibitors (Hsp90i)¹⁷. Likewise, histone deacetylase inhibitors, including FDA-approved SAHA, are promising anti-cancer drugs whose actions involve hyperacetylation of histone and select non-histone targets including HDAC6, thus indirectly inhibiting Hsp90 (ref. 21).

The cytotoxicity of 17AAG/SAHA in mutp53 cancer cells, despite being pleiotropic drugs, is largely due to the destabilization of mutp53 protein via Hsp90/HDAC6 inhibition^{11,12}. Moreover, owing to complementary drug targets, 17AAG/SAHA treatment caused synergistic cytotoxicity in human breast-cancer cells compared with monotherapy¹¹. Likewise, 17AAG and SAHA synergized in T47D (p53L194F) xenografts *in vivo* (Extended Data Fig. 3). SAHA or 17AAG alone were effective in blocking tumour growth of parental MDA231 (p53R280K) cells, but lost their efficacy when excess ectopic mutp53 was present. Only the combination of both drugs overcame this block (Fig. 3a). To translate the genetic proof-of-principle results from floxQ/− mice (Figs 1, 2) towards clinical application, we performed long-term treatments with 17DMAG+SAHA in mutp53R172H (H) mice³. Starting at 8 weeks when most H/H mice exhibited early intrathymic T-lymphoma (Fig. 3b), H/H and p53-null mice were treated lifelong with 17DMAG+SAHA versus vehicles. Strikingly, HSP90 inhibition benefited only H/H but not p53-null mice, extending their overall survival from a median 140 to 182 days ($P < 0.001$, Fig. 3c). Moreover, drug treatment improved survival of H/H mice beyond that of p53-null mice (Fig. 3d). For T-lymphoma-specific survival, again only H/H but not p53-null controls benefited from 17DMAG+SAHA treatment ($P < 0.001$, Fig. 3e). These data strongly support the idea that tumours expressing mutp53 depend on its presence and fundamentally differ in their oncogenic wiring from p53-null tumours. Moreover, the mechanism of action of these pleiotropic HSP90 inhibitors is mutp53-dependent.

Analysis of vehicle- and drug-treated thymic tissues at death for evidence of drug activity showed HSP90*α* activity (marked by transient induction of HSF1 target Hsp70 (ref. 22)), but this did not significantly prevent tumours in p53-null mice (Fig. 3f). In contrast, in H/H mice, when the drug worked (Hsp70 induced, mutp53 undetectable) it apparently prevented T-lymphoma formation in every case since we found only normal thymic tissues (Fig. 3f, 'p53H/H normal thymus' mice 7–14, plus 12 mice analysed by autopsy). In contrast, in all six T-lymphomas that arose despite drug treatment, the drugs failed to

inhibit HSP90 (Hsp70 not induced) (Fig. 3f, 6/26 (23%) escapers). Mechanistically, NQO1 (NAD(P)H:quinone oxido-reductase) is required to activate 17DMAG by metabolizing its quinone moiety, and NQO1 downregulation is a predictive biomarker for resistance to this ansamycin class of Hsp90i^{23,24}. Indeed, we observed reduced *NQO1* expression in escaper versus responder T-lymphomas (Fig. 3g). In sum, HSP90 inhibition prevented the development of T-lymphomas in *mutp53* mice.

To reduce complexity and dosing of the combinatorial anti-HSP90 treatment, we eliminated HDAC inhibition and replaced 17DMAG with the highly potent synthetic Hsp90i ganetespib. Ganetespib has demonstrated a favourable safety profile in more than 1,200 cancer patients treated so far, and is currently being evaluated in 11 clinical trials including a phase 3 lung cancer trial (www.syntapharma.com)^{25,26}. On a molar basis, ganetespib is >50-fold more potent than 17AAG in degrading mutp53 and killing mutp53 cancer cells (Extended Data Fig. 4a–c). Moreover, ganetespib destabilized mutp53 but not wild-type p53 in human tumour cells and xenografts (Extended Data Fig. 4d–f). Ganetespib also induced markedly more apoptosis in Q/– versus p53-null T-lymphoma cells (Extended Data Fig. 4g, h). Likewise, ganetespib strongly suppressed tumour growth in allografts of Q/– T- and B-lymphomas treated with prophylactic and therapeutic protocols (Fig. 4a, b and Extended Data Fig. 4i, j). Notably, this *in vivo* drug effect, marked by Hsp70 induction, correlated perfectly with mutp53 destabilization and induction of apoptosis within 24 h upon repeated dosing (Fig. 4c, d). Moreover, ganetespib synergized with conventional chemotherapy such as cyclophosphamide (Fig. 4e and Extended Data Fig. 4k).

Encouraged by these data, we performed long-term treatments of early disease (8 weeks) H/H and Q/− (Fig. 2f, 3b) versus their respective p53^{−/−} controls (note, they differ in genetic background^{23,27,28}) with one weekly dose of ganetespib (50 mg per kg) or vehicle lifelong. Strikingly, ganetespib benefited only mutp53 mice, representing a structural mutant (H allele) and a DNA-contact mutant (Q allele), but not p53-null mice, extending median overall survival of H/H mice from 157 to 233 days (48%, Fig. 4f) and Q/− mice from 146 to 232 days (59%, Fig. 4g). Q/− T-lymphoma-specific survival was also greatly extended (Fig. 4h). Importantly, clinically advanced autochthonous tumours in Q/− mice responded to ganetespib with regression or

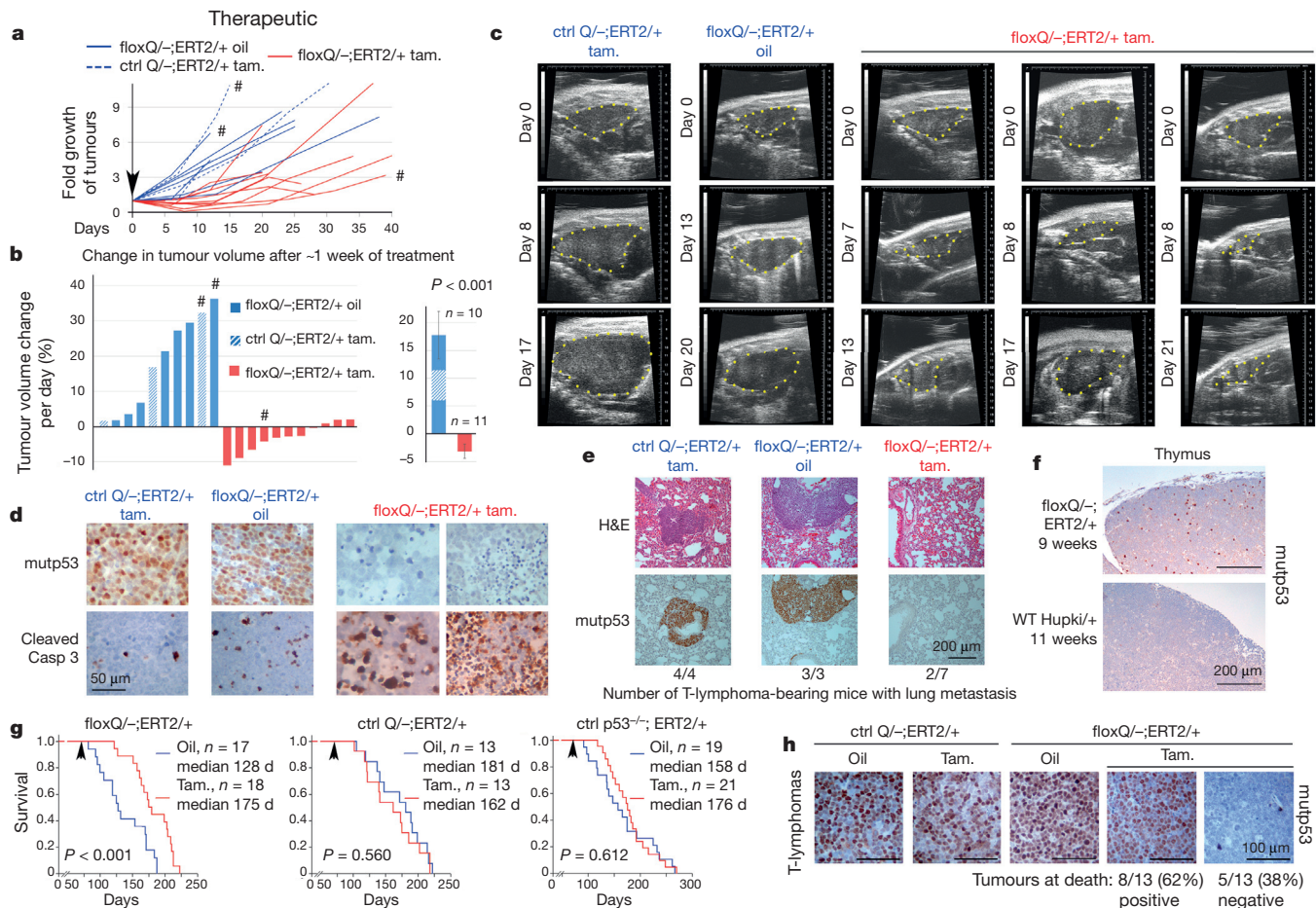


Figure 2 | Mutp53 ablation in floxQ^{-/-} mice causes autochthonous tumour regression or stagnation and extends survival. **a**, Fold growth over time until endpoint of clinically advanced tumours in floxQ^{-/-};ERT2^{+/+} mice imaged by ultrasound and treated with tamoxifen or oil (beginning at day 0, arrowhead); normalized to initial tumour size. **b**, Daily growth rates of individual tumours during the first 5–12 days of tamoxifen/oil treatment (left), and mean \pm s.e.m. of all tumours (right). Unpaired two-tailed Student's *t*-test. #, sarcomas; all others are T-lymphomas. **c**, Representative examples of sagittal ultrasound images of T-lymphomas in tamoxifen-treated floxQ^{-/-};ERT2^{+/+} and control mice. **d**, Genetic ablation of mutp53 in autochthonous tumours induces

stagnation (Fig. 4i). H/H mice, whose lifespan is normally identical to p53-null mice⁷, responded to ganetespib with improved survival beyond p53-null controls (Fig. 4j), whereas Q^{-/-} mice, whose lifespan is significantly shorter than p53-nulls² (Extended Data Fig. 1d), showed improved survival now resembling p53-nulls (Extended Data Fig. 4l). The fact that ganetespib and genetic mutp53 removal (Extended Data Fig. 2d) both cause floxQ^{-/-} survival to overlap with p53-null mice supports the idea that ganetespib targets mutp53. However, ganetespib improved survival more efficiently than genetic mutp53 ablation (Extended Data Fig. 4m), probably because of persistent destabilization of mutp53 compared with incomplete genetic ablation at 10 weeks. Other tumour-promoting Hsp90 clients might also be inhibited by ganetespib, contributing to the difference. Overall, these data indicate that the effect of ganetespib is largely due to targeting mutp53, as it had no effect in p53-null mice (Fig. 4f, g). Moreover, consistent with superior potency of ganetespib versus 17DMAG+SAHA, weekly ganetespib monotherapy improved overall survival more efficiently than near-daily 17DMAG+SAHA therapy (Extended Data Fig. 4n).

Ganetespib treatment also correlated with absence of T-lymphomas in two of two Q^{-/-} sibling pairs, while dimethylsulphoxide (DMSO)-treated siblings exhibited large T-lymphomas (Fig. 4k). Finally, thymic tissues of Q^{-/-} mice analysed at death again showed that when HSP90

apoptosis. Immunostaining for mutp53 and cleaved caspase 3 in representative control and mutp53-ablated T-lymphomas. **e**, Lung metastasis in these mice by haematoxylin and eosin (H&E) and p53 immunostaining. **f**, Organ-confined disease in young floxQ^{-/-} mice indicated by p53 immunostaining of malignant thymic cell clones. WT, wild type. **g**, Kaplan-Meier analysis and log rank statistics comparing cancer-related overall survival of floxQ^{-/-};ERT2^{+/+} versus Q^{-/-};ERT2^{+/+} and p53^{-/-};ERT2^{+/+} mice. Animals were treated once (arrow) at 10 weeks with tamoxifen or oil for 5 consecutive days. **h**, p53 immunostaining at endpoint (death) of representative T-lymphomas (see also Extended Data Fig. 2f).

inhibition was achieved (Hsp70 induced), mutp53 T-lymphomagenesis was prevented (Fig. 4l), while all five T-lymphomas that arose despite ganetespib again showed no evidence of drug activity (no Hsp70 induction) (Fig. 4l, 5/23 (22%) escapers). In colorectal cancers we recently identified a major resistance mechanism for ganetespib (but not other classes of Hsp90i) via upregulation of the drug-conjugating enzyme UGT1A in tumour cells, causing drug efflux and tumour protection²⁹. However, in mutp53 T-lymphomas, UGT1A (messenger RNA and protein) did not correlate with ganetespib resistance, necessitating future analysis to identify alternative mechanisms. Overall, these data indicate tumour dependence on high levels of mutp53 and demonstrate mutp53-dependent drug effects.

Missense mutp53 is highly expressed in 40–50% of all human tumours. One therapeutic strategy considered for nearly two decades is to develop small compounds capable of restoring the lost wild-type function of mutp53 proteins³⁰. In the face of this daunting and still elusive goal, the results presented here show that eliminating stabilized mutp53 proteins can have positive therapeutic effects *in vivo*, even in the absence of wild-type p53, since such tumours show exploitable dependence on their gain-of-function. Targeting HSP90 represents the first and currently only viable clinical strategy to achieve this goal.

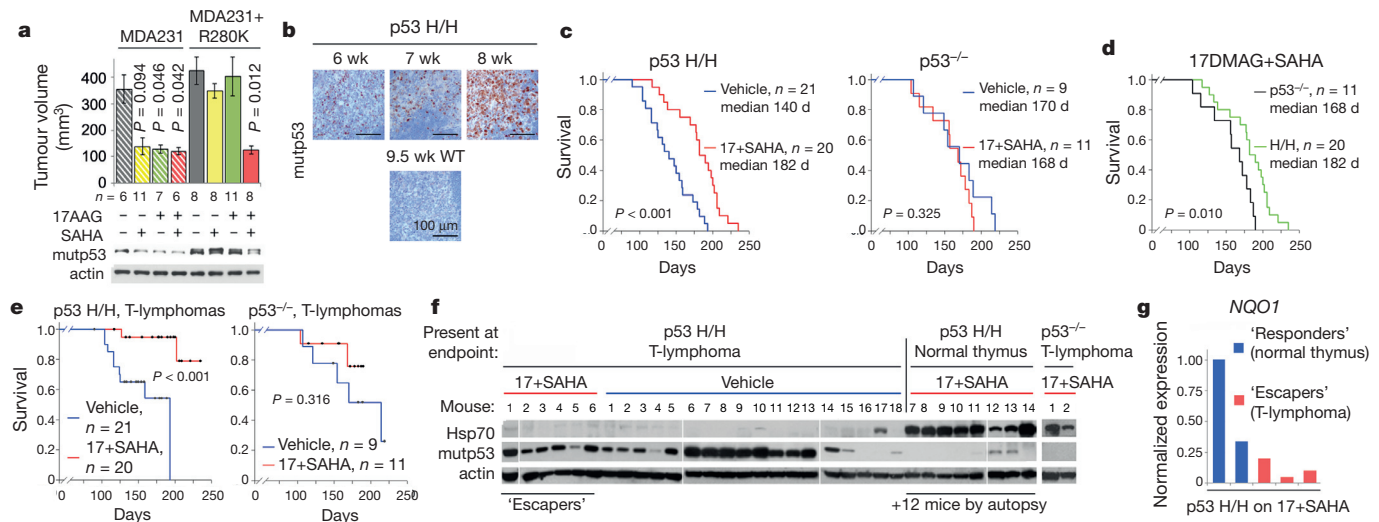


Figure 3 | Pharmacological inhibition of the mutp53 stabilizing HSP90/HDAC6 axis with 17DMAG+SAHA prolongs survival of H/H mice in a mutp53-dependent manner. **a**, 17AAG and SAHA synergize in degrading mutp53 and suppressing growth of subcutaneous xenografts of MDA231 cells expressing excess ectopic p53R280K. Mean \pm s.e.m.; *n*, number of xenografts, unpaired one-tailed Student's *t*-test. **b**, Organ-confined disease in young H/H mice, indicated by p53 immunostaining of malignant thymic cell clones.

c–e, cancer-related overall survival (**c**, **d**) and T-lymphoma-specific survival (**e**) of 17DMAG+SAHA-treated mutp53 H/H and p53-null mice, Kaplan–Meier analyses, log rank statistics. **f**, Immunoblot of p53 and Hsp70 in thymic tissues at endpoint from mice shown in **e**; each lane represents a different vehicle- or drug-treated mouse. **g**, Real-time PCR analysis of *NQO1* in representative responder and escaper H/H T-lymphomas from **f**.

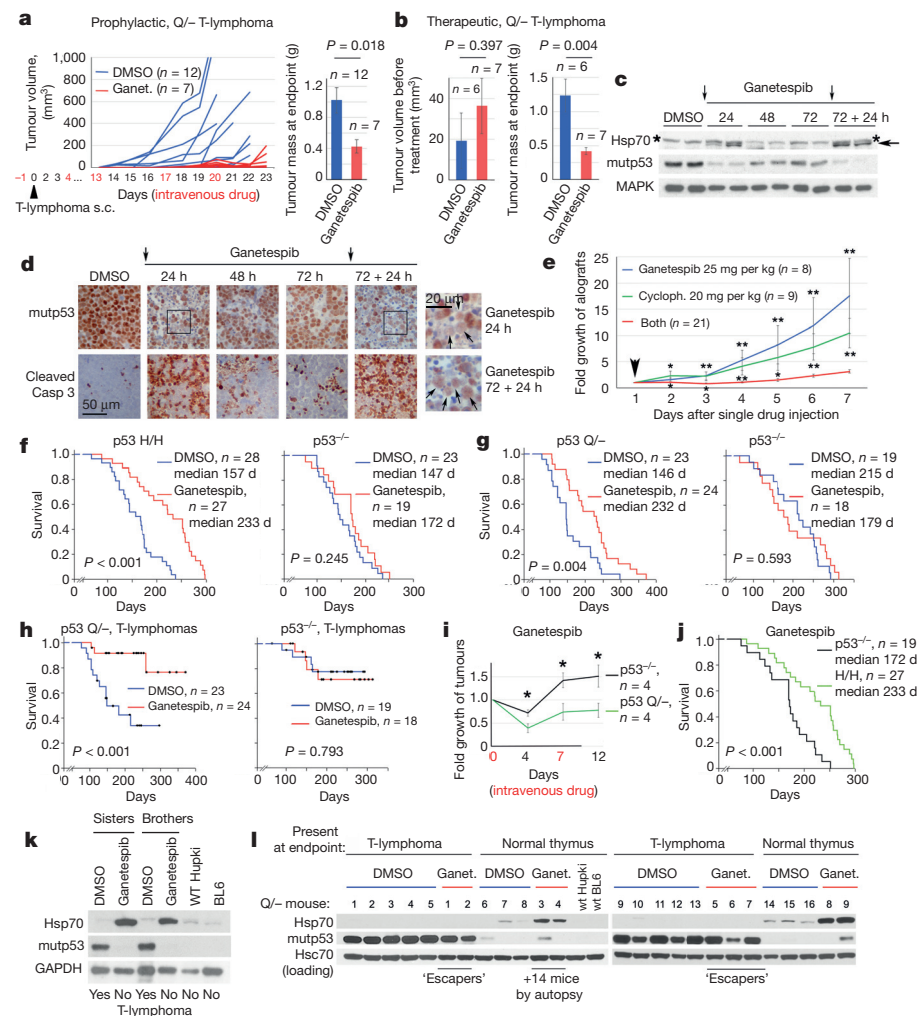


Figure 4 | Treatment of H/H and Q/- mice with ganetespib suppresses tumour growth and extends survival in a mutp53-dependent manner. **a**, **b**, Prophylactic (**a**) and therapeutic (**b**) treatment of subcutaneous Q/- allografts with ganetespib (ganet.) or vehicle (DMSO). **a**, Arrowhead, tumour cell injection. **b**, Initial allograft volume and mass at endpoint, mean \pm s.e.m., unpaired two-tailed Student's *t*-test; *n*, number of allografts. **c**, **d**, Time course of mutp53 levels and apoptosis (cleaved caspase 3) in ganetespib-treated (arrows) Q/- subcutaneous T-lymphoma allografts, analysed by immunoblot (**c**) and immunostaining (**d**). Asterisks mark non-specific bands (**c**). Enlarged panels after the first and second doses (**d**, far right) immunostained for p53 (arrows), dead cells stain blue. **e**, Growth of subcutaneous Q/- T-lymphoma allografts treated once (arrow) with ganetespib or cyclophosphamide alone or in combination. Mean \pm s.e.m., unpaired two-tailed Student's *t*-test; *n*, number of allografts, **P* < 0.05, ***P* < 0.01 (single drugs versus combination). **f–h**, Kaplan–Meier analysis and log rank statistics of cancer-related overall survival (**f**, **g**) and T-lymphoma-specific survival (**h**) of mutp53 H/H, Q/- and their respective p53^{-/-} controls. **i**, Fold growth over time of clinically advanced Q/- and p53^{-/-} T-lymphomas imaged by ultrasound and treated with ganetespib. Mean \pm s.e.m., unpaired two-tailed Student's *t*-test; *n*, number of tumours, **P* < 0.05. **j**, Comparison of overall survival of ganetespib-treated mutp53 H/H and their respective p53^{-/-} controls, shown in **f**. **k**, **l**, Immunoblot analysis of thymic tissues from same-litter siblings (**k**) and from Q/- mice from **h** at endpoint. Each lane represents individual DMSO- or ganetespib-treated mice.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 August 2014; accepted 20 March 2015.

Published online 25 May 2015.

- Doyle, B. *et al.* p53 mutation and loss have different effects on tumorigenesis in a novel mouse model of pleomorphic rhabdomyosarcoma. *J. Pathol.* **222**, 129–137 (2010).
- Hanel, W. *et al.* Two hot spot mutant p53 mouse models display differential gain of function in tumorigenesis. *Cell Death Differ.* **20**, 898–909 (2013).
- Lang, G. A. *et al.* Gain of function of a p53 hot spot mutation in a mouse model of Li-Fraumeni syndrome. *Cell* **119**, 861–872 (2004).
- Morton, J. P. *et al.* Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proc. Natl Acad. Sci. USA* **107**, 246–251 (2010).
- Olive, K. P. *et al.* Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome. *Cell* **119**, 847–860 (2004).
- Brosh, R. & Rotter, V. When mutants gain new powers: news from the mutant p53 field. *Nature Rev. Cancer* **9**, 701–713 (2009).
- Terzian, T. *et al.* The inherent instability of mutant p53 is alleviated by Mdm2 or p16INK4a loss. *Genes Dev.* **22**, 1337–1344 (2008).
- Suh, Y. A. *et al.* Multiple stress signals activate mutant p53 *in vivo*. *Cancer Res.* **71**, 7168–7175 (2011).
- Blagosklonny, M. V., Toretsky, J., Bohen, S. & Neckers, L. Mutant conformation of p53 translated *in vitro* or *in vivo* requires functional HSP90. *Proc. Natl Acad. Sci. USA* **93**, 8379–8383 (1996).
- Whitesell, L., Sutphin, P. D., Pulcini, E. J., Martinez, J. D. & Cook, P. H. The physical association of multiple molecular chaperone proteins with mutant p53 is altered by geldanamycin, an hsp90-binding agent. *Mol. Cell. Biol.* **18**, 1517–1524 (1998).
- Li, D., Marchenko, N. D. & Moll, U. M. SAHA shows preferential cytotoxicity in mutant p53 cancer cells by destabilizing mutant p53 through inhibition of the HDAC6-Hsp90 chaperone axis. *Cell Death Differ.* **18**, 1904–1913 (2011).
- Li, D. *et al.* Functional inactivation of endogenous MDM2 and CHIP by HSP90 causes aberrant stabilization of mutant p53 in human cancer cells. *Mol. Cancer Res.* **9**, 577–588 (2011).
- Bougeard, G. *et al.* Molecular basis of the Li-Fraumeni syndrome: an update from the French LFS families. *J. Med. Genet.* **45**, 535–538 (2008).
- Zerdoumi, Y. *et al.* Drastic effect of germline TP53 missense mutations in Li-Fraumeni patients. *Hum. Mutat.* **34**, 453–461 (2013).
- Muller, P. A. & Vousden, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* **25**, 304–317 (2014).
- Weissmueller, S. *et al.* Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor beta signaling. *Cell* **157**, 382–394 (2014).
- Trepel, J., Mollapour, M., Giaccone, G. & Neckers, L. Targeting the dynamic HSP90 complex in cancer. *Nature Rev. Cancer* **10**, 537–549 (2010).
- Whitesell, L. & Lindquist, S. L. HSP90 and the chaperoning of cancer. *Nature Rev. Cancer* **5**, 761–772 (2005).
- Dai, C., Whitesell, L., Rogers, A. B. & Lindquist, S. Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. *Cell* **130**, 1005–1018 (2007).
- Min, J. N., Huang, L., Zimonjic, D. B., Moskophidis, D. & Mivechi, N. F. Selective suppression of lymphomas by functional loss of Hsf1 in a p53-deficient mouse model for spontaneous tumors. *Oncogene* **26**, 5086–5097 (2007).
- Slingerland, M., Guchelaar, H. J. & Gelderblom, H. Histone deacetylase inhibitors: an overview of the clinical studies in solid tumors. *Anticancer Drugs* **25**, 140–149 (2014).
- Zou, J., Guo, Y., Guettouche, T., Smith, D. F. & Voellmy, R. Repression of heat shock transcription factor HSF1 activation by HSP90 (HSP90 complex) that forms a stress-sensitive complex with HSF1. *Cell* **94**, 471–480 (1998).
- Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
- Siegel, D., Yan, C. & Ross, D. NAD(P)H:quinone oxidoreductase 1 (NQO1) in the sensitivity and resistance to antitumor quinones. *Biochem. Pharmacol.* **83**, 1033–1040 (2012).
- Proia, D. A. & Bates, R. C. Ganetespib and HSP90: translating preclinical hypotheses into clinical promise. *Cancer Res.* **74**, 1294–1300 (2014).
- Socinski, M. A. *et al.* A multicenter phase II study of ganetespib monotherapy in patients with genotypically defined advanced non-small cell lung cancer. *Clin. Cancer Res.* **19**, 3068–3077 (2013).
- Harvey, M. *et al.* Genetic background alters the spectrum of tumors that develop in p53-deficient mice. *FASEB J.* **7**, 938–943 (1993).
- Jacks, T. *et al.* Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* **4**, 1–7 (1994).
- Landmann, H. *et al.* UDP glucuronosyltransferase 1A expression levels determine the response of colorectal cancer cells to the heat shock protein 90 inhibitor ganetespib. *Cell Death Dis.* **5**, e1411 (2014).
- Khoo, K. H., Verma, C. S. & Lane, D. P. Drugging the p53 pathway: understanding the route to clinical efficacy. *Nature Rev. Drug Discov.* **13**, 217–236 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to S. Schlick, D. Mott and P. Yang for technical assistance. This work was supported by National Cancer Institute award 1R01CA176647 and Deutsche Forschungsgemeinschaft award MO 1998/2-1.

Author Contributions U.M.M. conceived and designed the study, E.M.A., A.R.Y., D.L., S.X., R.S. and D.A.P. conducted experiments, E.M.A., A.R.Y., D.L. and M.D. analysed experiments, GL provided p53H/H and p53^{-/-} cohorts, D.A.P. provided ganetespib and U.M.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to U.M.M. (Ute.Moll@stonybrookmedicine.edu).

METHODS

Generation of floxQ;ERT2 mice. The floxed HUPKI (humanized p53 knock-in) p53R248Q allele (referred to as floxQ) was generated by introducing loxP sites flanking exons 2–10 of the p53 locus where mouse exons 4–9 were replaced with human exons 4–9 containing a p53 R248Q mutation in exon 7, as described in Extended Data Fig. 1a–c. The p53^{floxQ/+} mice were crossed with p53^{-/-} mice on a 129SvImJ/C57Bl6J mixed background as previously described² to generate p53^{floxQ/-} animals. The p53^{floxQ/-} animals were then intercrossed with p53^{Q/Q} animals (previously described in ref. 2) to generate p53^{floxQ/Q} mice on a 129SvImJ/C57Bl6J background. Rosa26^{CreERT2/CreERT2} mice (referred to as ERT2) on a pure C57Bl6J background³¹ were purchased from The Jackson Laboratory (stock 008463) and intercrossed with the mixed p53^{-/-} animals to generate p53^{-/-};Rosa26^{ERT2/ERT2} mice. The p53^{-/-};Rosa26^{ERT2/ERT2} mice were intercrossed with p53^{floxQ/Q} or p53^{-/-} to generate the p53^{floxQ/Q};Rosa26^{ERT2/+}, p53^{Q/Q};Rosa26^{ERT2/+} and p53^{-/-};Rosa26^{ERT2/+} cohorts.

Mutant p53 (mutp53) knock-in mice for long-term drug studies. The p53^{R172H/R172H} mice (referred to as H/H) and their corresponding p53^{-/-} controls on a pure C57Bl6J background were previously described^{3,27}. HUPKI p53^{R248Q/-} mice (referred to as Q/-) and their corresponding p53^{-/-} controls on a 129SvImJ/C57Bl6J mixed background were also previously described^{2,28}. Note that two different p53^{-/-} control strains were used as corresponding genetic background-matching controls for H/H mice (pure C57Bl6J background) versus Q/- mice (mixed 129SvImJ/C57Bl6J background). Mice were monitored twice a week and euthanized when animals appeared moribund or the largest palpable tumour reached 2 cm³ or ulcerated. Males and females were used at 1:1 ratio (except for Nu/Nu mice, only males were used). According to Stony Brook University Institutional Animal Care and Use Committee (IACUC) guidelines and our approved protocol, the maximum tumour burden per mouse was 4 cm³, which was strictly followed in all experiments. Time of euthanasia was considered endpoint. All animals were treated humanely and according to the guidelines by the IACUC of Stony Brook University.

Lymphoma cell isolation and allograft transplants. Primary T-lymphoma cells were isolated by mechanically mincing T-lymphoma tumours with a razor blade for 1 min, followed by 5 min incubation with 5 ml red blood cell lysis buffer at 22°C. Lymphoma medium (5 ml) (1:1 DMEM/IMDM supplemented with 10% FBS, L-glutamine, penicillin/streptomycin, antibiotic/antimycotic, β-mercaptoethanol; all from Gibco) was added and cells were passed through a 40 μm nylon cell strainer (Falcon) to remove cell clumps. Cells were spun down and resuspended in lymphoma medium. Primary B-lymphoma cells were isolated similarly, except for the addition of 100 μl of 3 M sodium acetate during mincing to prevent blood coagulation and a second round of red blood cell lysis buffer/lymphoma medium/spinning. Freshly isolated live B- and T-lymphoma cells were counted and either frozen or immediately used for *in vitro* studies or transplantations (allografts). For subcutaneous transplantation, 6- to 7-week-old Nu/Nu males (Harlan, strain Hsd:Athymic Nude-Foxn1^{nu}) were injected into four dorsal sites with 10⁵ lymphoma cells per site suspended in 3:1 PBS/Matrigel (BD Biosciences). For intravenous tail vein injections, 10⁵ cells in PBS were injected per Nu/Nu recipient.

Removal of the conditional floxQ allele. To remove the floxQ allele in cells *in vitro*, T-lymphoma cells (15 × 10⁶ live cells per 6 cm dish) grown in lymphoma medium were treated with 1 μM 4OHT (Sigma H7904) in EtOH for the times indicated in Extended Data Fig. 1e.

To remove the floxQ allele in T-lymphoma allografts, randomly assigned mice were injected intraperitoneally with 100 mg per kg or 150 mg per kg tamoxifen (Sigma T5648) in corn oil, or with corn oil alone, as indicated in Fig. 1 and Extended Data Fig. 1. For the prophylactic protocol typically nine daily intraperitoneal injections beginning 2 days before lymphoma cell transplantation were given. For the therapeutic protocol five to seven daily intraperitoneal injections were given after at least two out of four palpable tumours had appeared.

For autochthonous tumours, to remove the floxQ allele in mice with early organ-confined disease, randomly assigned 10-week-old animals were injected intraperitoneally with 100 mg per kg tamoxifen in corn oil, or corn oil alone, daily for 5 consecutive days (Fig. 2f–h and Extended Data Fig. 2c–f). As a pre-set criterion, animals that died from non-cancer-related causes (the minority) were excluded from subsequent analyses. To treat clinically advanced tumours (Fig. 2a–e), animals were monitored weekly by ultrasound imaging using the Visual Sonics Vevo 770 High-Resolution Imaging System. When tumours had reached 150 mm³ on average (ranging from 86 to 218 mm³), mice were injected with 150 mg per kg tamoxifen daily for 7 days, followed by 100 mg per kg three times a week, or with oil, until they needed to be killed (endpoint). To follow the response *in vivo*, tumours were imaged weekly by ultrasound and tumour volume was calculated using Vevo 770 V3.0.0 software.

Hsp90 inhibitor studies. HSP90 is crucial for cancer cell survival³². Cultured cells were treated with 2.5 μM or 5 μM 17-allylamino-17-demethoxygeldanamycin (17AAG, LC Laboratories, A-6880) or 5–500 nM ganetespib in DMSO for 24 h or 48 h. For xenografts, human cancer cells were subcutaneously injected into nude mice (0.5 Mio cells per site in 100 μl Matrigel, four sites per mouse). Tumours became visible around day 10. For Fig. 3a and Extended Data Fig. 3, treatment of 10 mg per kg 17AAG and/or 25 mg per kg SAHA (vorinostat/suberoylanilide hydroxamic acid, LC Laboratories, V-8477) in DMSO or vehicle control was started at day 24 every other day, with endpoint at day 44 or earlier in the few cases where tumours became ulcerated.

For 17DMAG+SAHA treatments of H/H and corresponding p53^{-/-} mice, 17DMAG (17-dimethylaminoethylamino-17-demethoxygeldanamycin, LC Laboratories D-3440), a hydrophilic derivative of 17AAG³³, was dissolved in 5% (w/v) hydroxypropyl-β-cyclodextrin (HOP-β-CD) (Sigma 332607), and SAHA (LC Laboratories V-8477) was dissolved in a 5 M equivalent of HOP-β-CD as previously described³⁴. Equal volumes of 17DMAG and SAHA (or both vehicle solutions) were combined and frozen in aliquots until use. Starting at 8 weeks of age, randomly assigned animals received 100 μl intraperitoneal injections of 10 mg per kg 17DMAG plus 25 mg per kg SAHA, or 100 μl HOP-β-CD vehicle, for 5 days a week until endpoint (death). As a pre-set criterion, animals that died from non-cancer-related causes (the minority) were excluded from subsequent analyses.

For ganetespib treatments of H/H and Q/- mice and their corresponding p53^{-/-} control cohorts, ganetespib (aka STA-9090, Synta Pharmaceuticals) was dissolved in 10% DMSO/18% cremophor/3.6% dextrose ('10/18 DRD') as previously described³⁵. For reference, rodents safely tolerate doses of ganetespib of up to 150 mg per kg once weekly or 75 mg per kg twice weekly³⁶. Randomly assigned animals received freshly prepared ganetespib or the corresponding volume of 10/18 DRD vehicle by tail vein injection as follows: 50 mg per kg ganetespib once a week starting at 8 weeks of age until death for the treatment of autochthonous tumours (including the sibling analysis in Fig. 4j); 75 mg per kg ganetespib twice a week for the treatment of T- and B-lymphoma allografts (beginning 1 day before lymphoma transplantation in the prophylactic protocol, or beginning when at least two out of four palpable tumours per mouse had appeared in the therapeutic protocol); 150 mg per kg of ganetespib once or twice for acute high-dose treatment (Fig. 4c, d). As a pre-set criterion, animals that died from non-cancer-related causes (the minority) were excluded from subsequent analyses. For the sibling analysis (Fig. 4j), the sisters were killed at 90 days and the brothers at 148 days, each 24 h after their last treatment, and their thymic tissues were subjected to immunoblot analysis. Cyclophosphamide was dissolved in PBS and injected intraperitoneal at indicated doses.

Human cancer cell lines. Breast cancer MDA231 (p53R280K, HTB-26), MDA468 (p53R273K, HTB-132), T47D (p53L194F, HTB-133), SKBR3 (p53R175H, HTB-30), prostate cancer DU145 (p53P223L and p53V274F, HTB-81), non-small-cell lung cancer H1975 (p53R273H, CRL-5908) and ovarian cancer cell line ES2 (p53S241F, CRL-1978) were purchased from the American Type Culture Collection. The ovarian cancer lines EFO21 (p53C124R, ACC-235) and COLO704 (p53wt, ACC-198) were purchased from Deutsche Sammlung von Mikroorganismen und Zellkulturen. The ovarian cancer cell line COV434 (p53wt) was purchased from Sigma-Aldrich (07071909). The ovarian cancer cell line HOC7 (p53C275F) was a gift from N. Concin. Paired parental MDA231 and MDA231+R280K cells stably overexpressing pcDNA3 vector or pcDNA3-p53R280K plasmids, respectively, were generated by transfection and selection. All cells were cultured in DMEM/10% FCS and tested negative for mycoplasma. No further authentication of cell lines was performed.

Cell death. Cell death was determined by CellTiter-Blue Cell Viability Assay (Promega G808A) according to the manufacturer's instructions, in six-well plate format with five Mio lymphoma cells per well or 150,000 human cancer cells per well. Cells were treated as indicated and fluorescence was detected by FILTERmax F5 (Molecular Devices). Trypan Blue exclusion assays were done by direct manual counting of transparent (live) cells in a haemocytometer and normalizing counts to vehicle control.

Immunohistochemistry and immunoblot analysis. For immunohistochemical analysis, tumours were fixed in formalin, embedded in paraffin and sectioned (5 μm). Slides were deparaffinized and boiled in citrate buffer (10 mM, pH 6.0, 35 min) for antigen retrieval, blocked in 10% goat serum and incubated with primary antibodies (mutp53, Santa Cruz FL393, sc-6243, 1:500; cleaved caspase 3, Cell Signaling 9661, 1:500; phosphor-S28 histone H3, Abcam ab32388, 1:300) for 2 h at room temperature. After PBS washing, slides were incubated with biotinylated secondary antibody and HRP-streptavidin using the Histostain SP Broad Spectrum kit (Invitrogen 959943B), stained with DAB substrate with haematoxylin counterstain and coverslipped. Immunoblots were performed using 20 μg of total protein extract from tumours or tumour cells with the following

antibodies: mutp53 (FL-393, Santa Cruz, sc-6243; DOI1, Santa Cruz sc-126), Hsp70 (Enzo C92F3A-5), PARP (Cell Signaling 9542), cleaved PARP (Cell Signaling 9541), cleaved caspase 3 (Cell Signaling 9661), Chk1 (Cell Signaling 2360), CDK1 (Millipore MAB8878), MAPK (Millipore 05-157), actin (Thermo Scientific MS-1295-P0), Hsc70 (Santa Cruz, sc-7298), GAPDH (Santa Cruz, sc-25778).

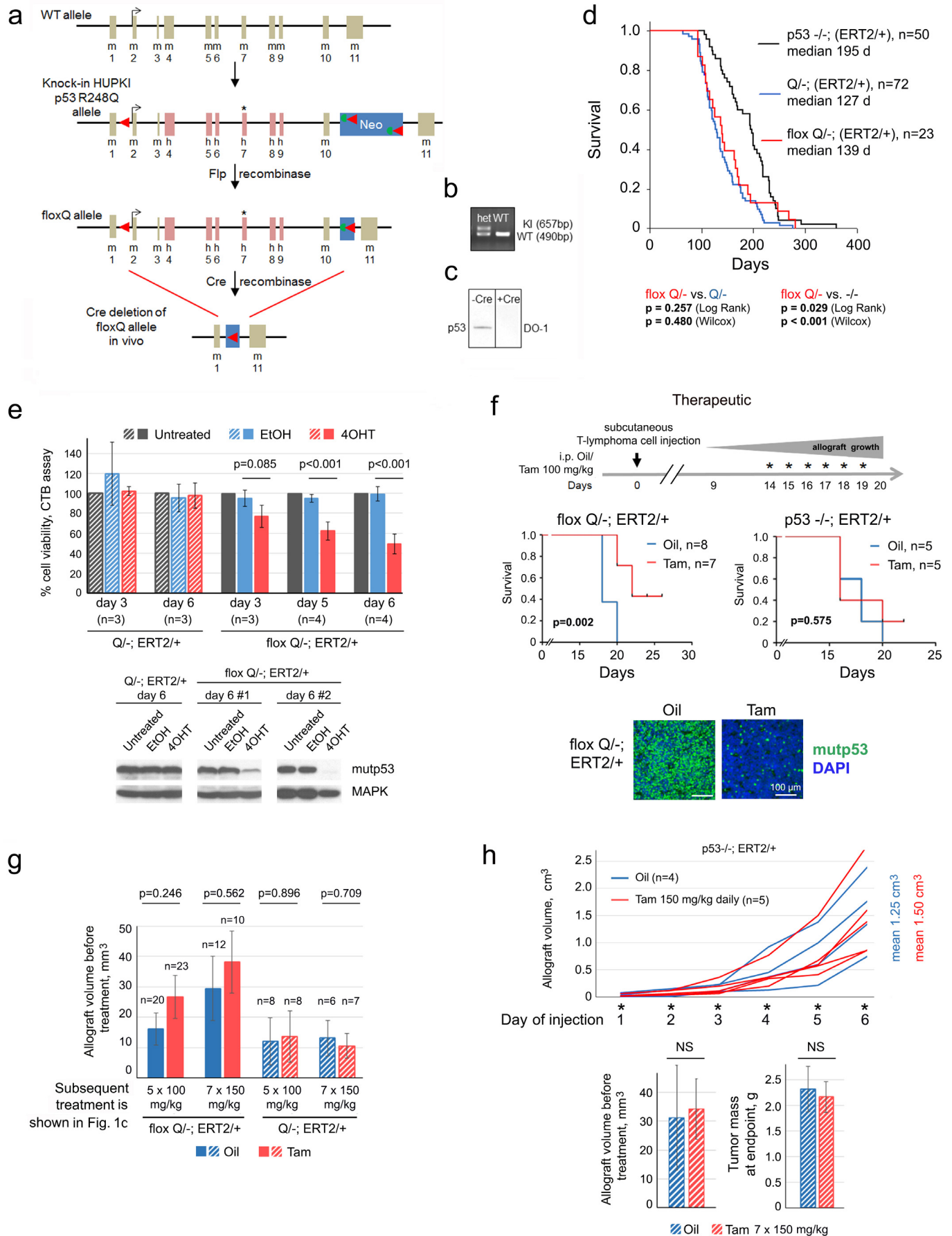
Quantitative PCR analysis. Total RNA was isolated using TRIzol reagent (Invitrogen) and 2 µg was reverse-transcribed with random primers and SuperScript II Reverse Transcriptase (Invitrogen 18064-014). Real-time PCR was performed in duplicates with QuantiTect SYBR Green Mix (Qiagen 204143) using the MJ Research DNA Engine Opticon 2 machine. Primers for mouse *NQO1* were: 5' TGGCCGAACACAAGAAGCTG 3' (forward), 5' GCTACGAGCACTCTCTCAAACC 3' (reverse). *NQO1* expression was normalized to the housekeeping gene *HPRT*.

Statistics. Statistical tests with appropriate underlying assumptions on data distribution and variance characteristics were used. Kaplan–Meier analyses with log rank statistics for animal survival curves were generated using SigmaPlot-Systat software version 11. Note: in Figs 3e and 4h, dots represent censored animals, namely mice dead owing to cancer types other than T-lymphomas. Each line ends

when the last animal dies (if caused by T-lymphoma, the line goes down to zero; if caused by another cancer type, the animal is censored and the line stops with a dot).

An unpaired two-tailed Student's *t*-test was used to analyse tumour measurements. No statistical method was used to predetermine sample size. No blinding was used. All *n* numbers indicate biological replicates, unless indicated otherwise.

31. Ventura, A. *et al.* Restoration of p53 function leads to tumour regression in vivo. *Nature* **445**, 661–665 (2007).
32. Dobbstein, M. & Moll, U. Targeting tumour-supportive cellular machineries in anticancer drug development. *Nature Rev. Drug Discov.* **13**, 179–196 (2014).
33. Gorska, M. *et al.* Geldanamycin and its derivatives as Hsp90 inhibitors. *Front. Biosci.* **17**, 2269–2277 (2012).
34. Hrzenjak, A. *et al.* Histone deacetylase inhibitor vorinostat suppresses the growth of uterine sarcomas *in vitro* and *in vivo*. *Mol. Cancer* **9**, 49 (2010).
35. Proia, D. A. *et al.* Multifaceted intervention by the Hsp90 inhibitor ganetespib (STA-9090) in cancer cells with activated JAK/STAT signaling. *PLoS ONE* **6**, e18552 (2011).
36. Ying, W. *et al.* Ganetespib, a unique triazolone-containing Hsp90 inhibitor, exhibits potent antitumor activity and a superior safety profile for cancer therapy. *Mol. Cancer Ther.* **11**, 475–484 (2012).

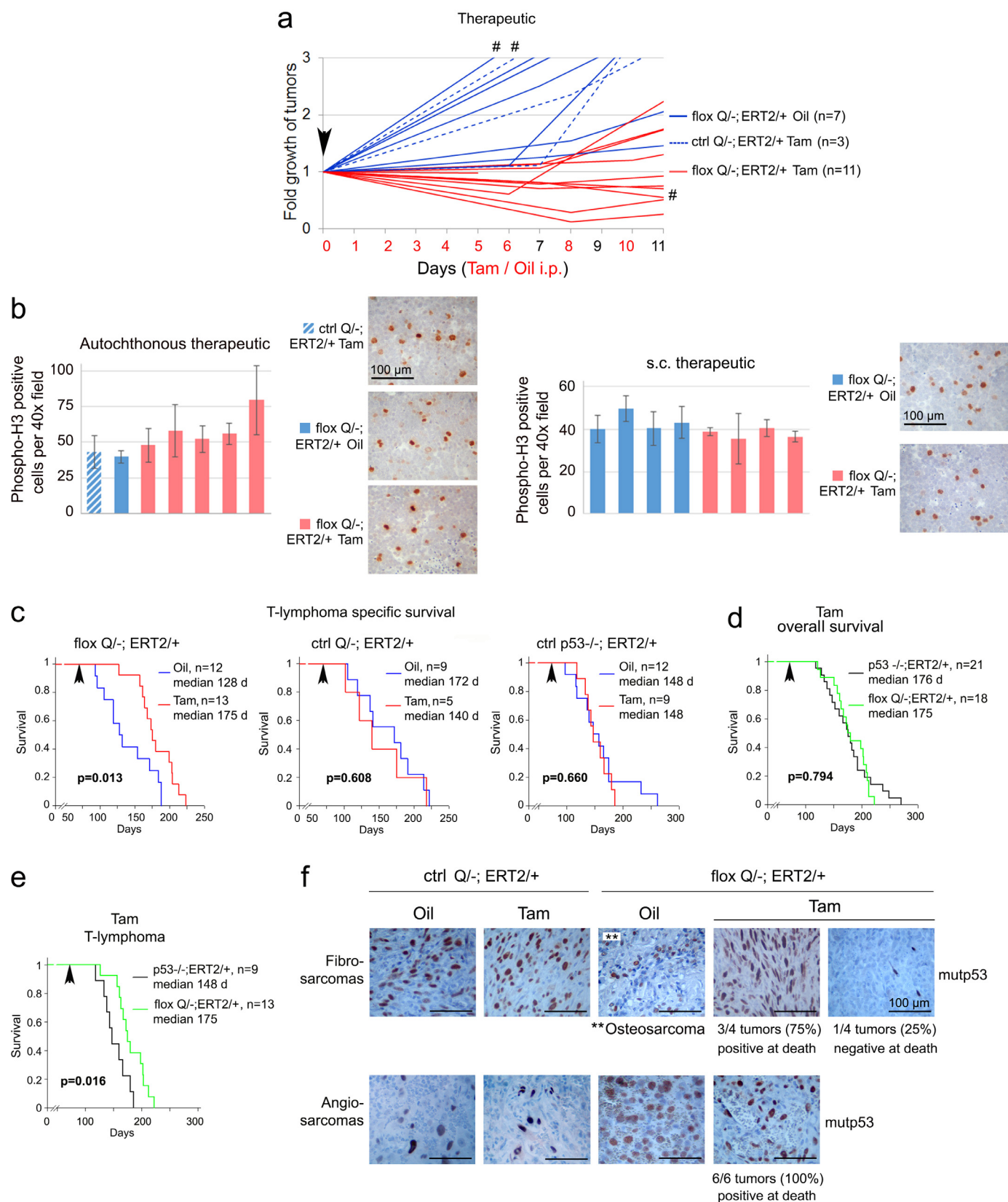


Extended Data Figure 1 | Generation and characterization of the

conditionally inactivatable p53 flox R248Q allele. **a**, Mouse exons 4–9 were replaced with human exons 4–9 (called HUPKI) containing a p53 R248Q mutation in exon 7 (marked by a star). Exons 2–10 were flanked with loxP sites in introns 1 and 10 (red arrows) to allow for Cre-targeted removal of the mutp53 allele upon addition of tamoxifen/4OHT. A deletable Neo selection box was flanked by FRT sites (green). Knock-in mice were mated with FLP mice to delete the Neo cassette *in vivo*, leaving behind the distal loxP site. The 'floxQ' allele thus has two loxP sites for subsequent Cre deletion. **b**, For genotyping, the Neo-deleted floxQ knock-in allele produces a 657-base-pair amplicon, in contrast to the 490-base-pair amplicon derived from the wtp53 allele. **c**, Normal mouse embryo fibroblasts (MEFs) from floxQ/– embryos, which as non-malignant cells express non-stabilized mutp53, were adenovirally infected with empty vector (–Cre) or Cre-expressing vector (+Cre). Cre-mediated deletion of the mutp53 allele was confirmed by immunoblot analysis. **d**, The floxQ and constitutive Q ('Q') alleles behave identically in all aspects of gain-of-function including overall survival and tumour spectrum (not shown). Both floxQ and Q mice predominantly develop aggressive T-lymphomas, with some additional B-lymphomas and sarcomas. Also, the RosaCreERT2 allele has no discernable impact (data not shown). Kaplan–Meier analysis comparing overall survival of floxQ/– (red), Q/– (blue) and p53^{–/–} (black) mice. Significance was assessed by log rank and Wilcoxon tests. **e**, Deletion of the mutp53 allele induces cell death *in vitro*. Viability of primary T- lymphoma cells freshly

harvested from floxQ/–;ERT2/+ mice ($n = 4$) and Q/–;ERT2/+ control mice ($n = 3$), untreated or treated once with 4OHT or vehicle (EtOH) in short-term culture for 3–6 days. CellTiter-Blue (CTB) assay, unpaired two-tailed Student's *t*-test; mean \pm s.e.m.; n , number of independent T-lymphomas. Bottom, corresponding immunoblots of representative T-lymphomas at day 6.

f, Deletion of mutp53 improves survival of host mice. Therapeutic protocol with primary floxQ/– T-lymphomas allotransplanted (black arrow on time axis) via subcutaneous injections into SCID mice. After visible tumours appeared, SCID mice were treated with daily intraperitoneal injections of oil or tamoxifen (star on time axis). Mice were killed when allowable endpoint size (1.5 cm³) was reached. Kaplan–Meier analysis, log rank test. Tamoxifen-induced allele deletion was strong but incomplete, shown by representative p53 immunofluorescence staining of tumours at endpoint (4',6-diamidino-2-phenylindole (DAPI) counterstain). **g**, Initial tumour volumes measured before treatment was started in the therapeutic protocol of the various groups shown in Fig. 1c. Unpaired two-tailed Student's *t*-test; mean \pm s.e.m.; n , number of allografts. **h**, Control for Fig. 1d. Therapeutic treatment of nude mice allografted with p53^{–/–};ERT2/+ T-lymphoma cells and treated with tamoxifen (150 mg per kg for 7 days) as indicated in the scheme in Fig. 1c (endpoint 2). No response to tamoxifen. Time course, initial allograft volume and tumour mass at endpoint. Unpaired two-tailed Student's *t*-test; mean \pm s.e.m.; n , number of allografts; NS, not significant.

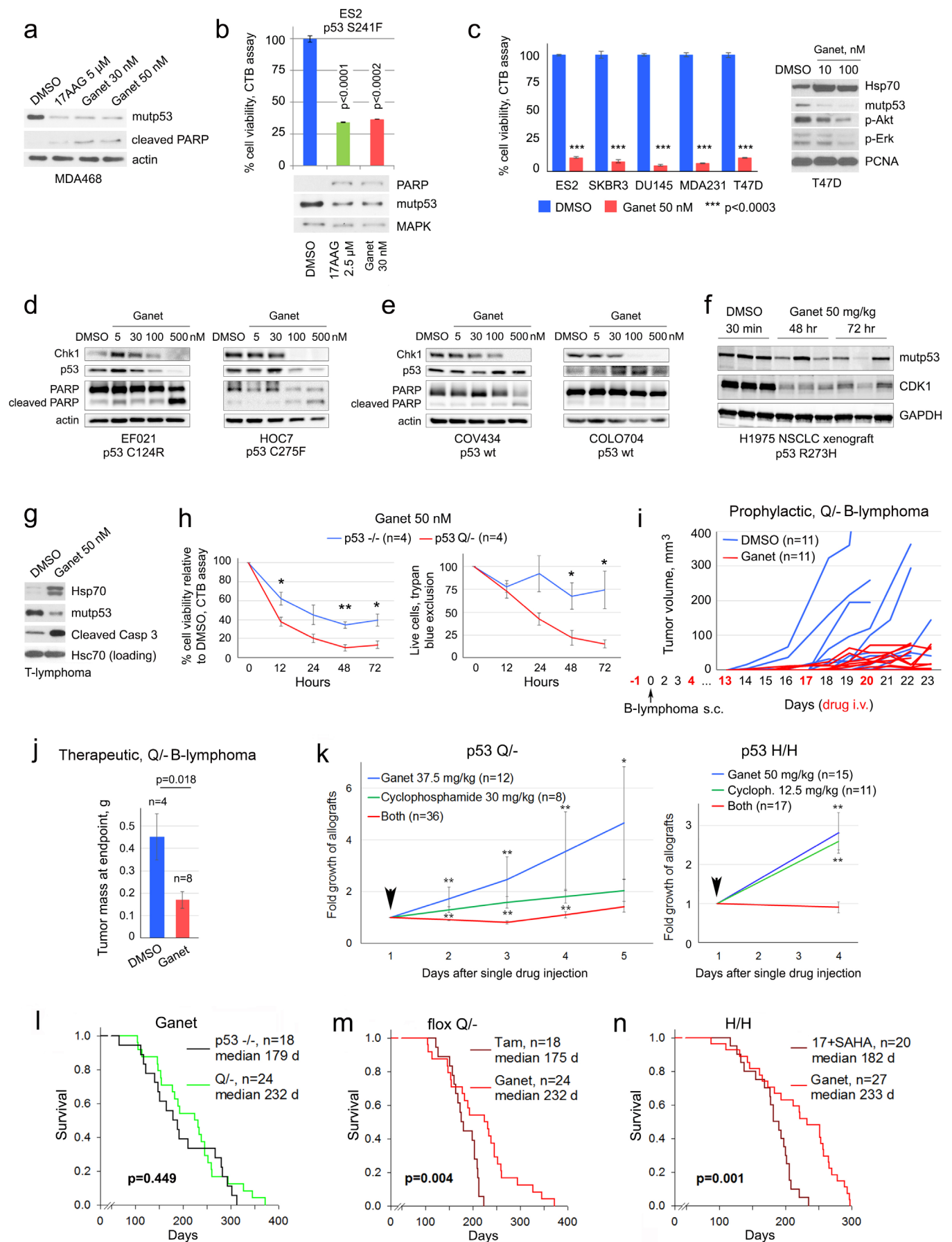


Extended Data Figure 2 | Mutp53 ablation in floxQ/− mice causes autochthonous tumour regression or stagnation and extends T-lymphoma-specific survival. **a**, Time course of imaged tumours, normalized to their initial tumour size (same as Fig. 2a but zoomed into the first 12 days of treatment). Stagnation or regression of floxQ/−;ERT2/+ tumours treated with tamoxifen, in contrast to treated control tumours (floxQ/−;ERT2/+ on oil and Q/−;ERT2/+ on tamoxifen) which grow robustly. Symbol #, sarcomas; all others are T-lymphomas. **b**, Similar mitotic index supports the idea that cell cycle arrest and senescence are not significantly affected upon genetic mutp53 ablation in autochthonous T-lymphomas (from Fig. 2a–c) and subcutaneous allografts (from Fig. 1c, d). Instead, apoptosis is the main mechanism of tumour regression/stagnation. Quantification of phospho-S28 histone H3 (pH3)-positive cells in individual autochthonous tumours or allografts. Five (left) or three (right) random $\times 40$ high-power fields (with no or only minimal apoptosis) were counted for each tumour. Mean \pm s.d. are plotted. Representative pH3 immunostainings are shown. **c–e**, Kaplan–Meier analyses

and log rank statistics comparing tamoxifen-treated T-lymphoma-specific survival (**c**, **e**) and overall (**d**) survival of floxQ/−;ERT2/+ mice versus constitutive Q/−;ERT2/+ and p53^{−/−};ERT2/+ control mice from Fig. 2g. Animals were treated once (arrow) at 10 weeks with oil or tamoxifen by intraperitoneal injections for 5 consecutive days. **f**, At endpoint (death), like T-lymphomas (Fig. 2h), most sarcomas in tamoxifen-treated floxQ/−;ERT2/+ mice are also entirely composed of p53-positive cells. This indicates strong selective pressure for mutp53-positive tumour cells in that the small minority of non-recombined malignant cells outcompeted the vast majority of recombined mutp53-deleted cells and with time took over the tumour mass, supporting tumour addiction to high levels of mutp53. Out of ten sarcomas, nine (90%) stained positive for p53 and only one (10%) was negative for p53. Immunostainings for p53 of representative fibro- and angiosarcomas are shown. Note, the blue cells in oil-treated osteosarcoma are normal stromal cells.



Extended Data Figure 3 | Synergistic action of 17AAG+SAHA in subcutaneous xenografts of mutp53-harboured T47D (p53 L194F) human breast cancer cells. Representative images of nude mice and their dissected tumours with one Mio cell injected per site.



Extended Data Figure 4 | Ganetespiib kills mutp53 human and mouse cells in a mutp53-dependent manner. **a–c**, On a molar basis, ganetespiib is >50-fold more potent than 17AAG in degrading mutp53 and killing human mutp53 cancer cells. MDA468 (p53R280K) (**a**) and T47D (p53L194F) (**c**) breast cancer cells, as well as ES2 (p53S241F) ovarian cancer cells (**b**), were seeded into six-well plates and treated for 24–48 h. After incubation, dead cells were washed off and total protein lysates from only live cells were immunoblotted as indicated. CTB assays on parallel cultures for cell viability show drug activity. **c**, SKBR3, (p53R175H) breast cancer cells; DU145, heterozygous (p53P223L/V274F) prostate cancer cells. Mean \pm s.e.m. of four (**b**) or three (**c**) technical replicas, unpaired two-tailed Student's *t*-test. p-Akt and p-Erk are also Hsp90 clients; cleaved PARP indicates activated apoptosis. **d–f**, Ganetespiib destabilizes mutp53 but not wtp53 in cultured human ovarian carcinoma cells EFO21 (p53C124R) and HOC7 (p53C275F) (**d**), wtp53 COV434 and COLO704 (**e**) and in human non-small-cell lung cancer xenografts H1975 (p53R273H) (**f**). Nude mice bearing tumour xenografts (each lane is an independent tumour) were treated with a single bolus of DMSO or ganetespiib (50 mg per kg intravenously), tumours were harvested at baseline (30 min), 48 h and 72 h and cells were lysed and immunoblotted as indicated (**f**). Chk1 and CDK1 are other Hsp90 clients; cleaved PARP indicates activated apoptosis. **g, h**, Ganetespiib decreases stabilized mutp53 levels in live Q/– T-lymphoma cells within 24 h, associated with induction of apoptosis. Freshly isolated live Q/– T-lymphoma cells were treated with DMSO or 50 nM ganetespiib for 24 h, followed by immunoblots as indicated, Hsp70 indicates drug activity, Hsc70 is the loading control (**g**). Death curves of freshly isolated Q/– and p53^{–/–} T-lymphoma cells treated with DMSO or 50 nM ganetespiib for the indicated times, CTB and trypan blue exclusion assays (**h**). All values are relative to DMSO treatment at the same time point. Mean \pm s.e.m., unpaired two-tailed Student's *t*-test, *n* = 4 independent isolates per genotype for every

time point, **P* < 0.05, ***P* < 0.01. **i, j**, Ganetespiib suppresses growth of subcutaneous allografts of Q/– B-lymphoma. Prophylactic protocol; treatment days are indicated in red, tumour cell injection is marked by arrow, time course of allograft growth; *n*, number of allografts (**i**), Therapeutic protocol (same as in Fig. 4b) and tumour mass at endpoint (**j**). Mean \pm s.e.m.; unpaired two-tailed Student's *t*-test; *n*, number of allografts. **k**, Ganetespiib yields synergistic anti-tumour effects in combination with cyclophosphamide. Subcutaneous allografts of Q/– and H/H T-lymphoma cells were treated once (arrow) with the indicated doses of ganetespiib or cyclophosphamide alone or in combination. The mean \pm s.e.m. allograft size for Q/– and H/H at the start of treatment was 274 \pm 36 mm³ and 323 \pm 44 mm³, respectively. Unpaired two-tailed Student's *t*-test; *n*, number of allografts. Each single drug is compared with the combination. **P* < 0.05, ***P* < 0.01. **l**, Comparison of ganetespiib treatment of floxQ/– versus corresponding p53^{–/–} control mice from Fig. 4g. The floxQ/– mice, which normally have a significantly shorter lifespan than p53-null littermates² (median 139 days versus 195 days, respectively, see also Extended Data Fig. 1d) respond to ganetespiib with significantly longer survival (right shift) and now resemble that of p53^{–/–} mice. Kaplan–Meier analysis, log rank statistics. **m, n**, Ganetespiib monotherapy once a week improves overall survival more efficiently than either genetic mutp53 ablation or 17DMAG+SAHA given five times a week. Comparison of Kaplan–Meier survival curves (log rank statistics) of tamoxifen-treated floxQ/– animals from Fig. 2g and ganetespiib-treated floxQ/– animals from Fig. 4g (**m**). Note, on the basis of their phenotypic identity (see Extended Data Fig. 1d), floxQ/– mice were used in Fig. 4g in lieu of Q/– to ensure direct comparability with tamoxifen treatment. Comparison of Kaplan–Meier survival curves (log rank statistics) of 17DMAG+SAHA-treated H/H animals from Fig. 3c and ganetespiib-treated H/H animals from Fig. 4f (**n**).

A noisy linear map underlies oscillations in cell size and gene expression in bacteria

Yu Tanouchi^{1*}, Anand Pai^{1*}, Heungwon Park^{2,3*}, Shuqiang Huang¹, Rumen Stamatov⁴, Nicolas E. Buchler^{2,3,5} & Lingchong You^{1,5}

During bacterial growth, a cell approximately doubles in size before division, after which it splits into two daughter cells. This process is subjected to the inherent perturbations of cellular noise^{1,2} and thus requires regulation for cell-size homeostasis. The mechanisms underlying the control and dynamics of cell size remain poorly understood owing to the difficulty in sizing individual bacteria over long periods of time in a high-throughput manner. Here we measure and analyse long-term, single-cell growth and division across different *Escherichia coli* strains and growth conditions³. We show that a subset of cells in a population exhibit transient oscillations in cell size with periods that stretch across several (more than ten) generations. Our analysis reveals that a simple law governing cell-size control—a noisy linear map—explains the origins of these cell-size oscillations across all strains. This noisy linear map implements a negative feedback on cell-size control: a cell with a larger initial size tends to divide earlier, whereas one with a smaller initial size tends to divide later. Combining simulations of cell growth and division with experimental data, we demonstrate that this noisy linear map generates transient oscillations, not just in cell size, but also in constitutive gene expression. Our work provides new insights into the dynamics of bacterial cell-size regulation with implications for the physiological processes involved.

We used a ‘mother machine’ microfluidic device³ and time-lapse microscopy to monitor long-term cell-size dynamics in *E. coli* at the single-cell level. The device enables the measurement of cell size and gene expression for hundreds of *E. coli* mother lineages over thousands of minutes and also allows continuous medium infusion to maintain balanced growth³. We first analysed temporal dynamics of the initial cell size (cell size at birth, or L_I) by computing its autocorrelation function (ACF). To our initial surprise, some lineages of mother cells (30–40%) exhibited long-term oscillations in cell size whose period could be longer than 10 generations. These oscillations were masked when averaged across the ensemble of all mother lineages (Fig. 1a–c). The periods of these oscillations were variable across lineages. For example, the period was about eight generations in one lineage (Fig. 1b) but about 16 generations in another (Fig. 1c).

How these long-term oscillations emerge, and why they have variable periods and only occur in a subset of cell lineages remain unclear. At a fundamental level, oscillations require negative feedback, and we attempted to address what might constitute this negative feedback in our system. We reasoned that negative feedback emerged from cell-size control: to maintain an average cell size over generations, a cell may sense its size and adjust either its growth rate (in biomass accumulation) during a cycle, the length of the cell cycle, or both. Such a control mechanism could provide the required negative feedback.

To test this notion, we measured various growth-related parameters (Fig. 1d) as a function of the initial cell size (Fig. 1e and Extended Data Fig. 1). We found that the final cell size (cell size before division, or L_F)

could be described, on average, by a linear function of the initial cell size, $L_F = aL_I + b$ (Fig. 1e). The slope of this linear function ($a = 0.871$) was <2 , which reflects negative-feedback control of cell size. We verified that this linear function also holds for different growth conditions (27 °C and 25 °C) and two other *E. coli* strains (MG1655 and B/r; data sets from a previous study³) (Extended Data Fig. 2). Our data showed that both the division ratio (R) and growth rate (μ) were relatively independent of initial size (Extended Data Fig. 1a, b). However, the doubling time (T) was negatively correlated with the initial size (Extended Data Fig. 1c), providing the basis for the negative-feedback control. That is, a mother cell with smaller initial size tended to grow for a longer duration than the average; a mother cell with a larger initial size tended to grow for a shorter duration than the average. Recent studies^{4–6} examining cell-size homeostasis in bacteria also reported this observation of modulation of division time.

On the basis of our observation of a linear relationship between L_I and L_F , we examined the extent to which a simple autoregressive model with noise might explain our experimentally observed data:

$$L_I(n+1) = (aL_I(n) + b + \eta)/2 \quad (1)$$

Here, $L_I(n)$ is the initial size at generation n , and η is Gaussian white noise, representing the scatter around the linear regression line in Fig. 1e. Using this ‘noisy linear map’ between the initial and final cell sizes, we numerically simulated the dynamics of L_I in 100 lineages,

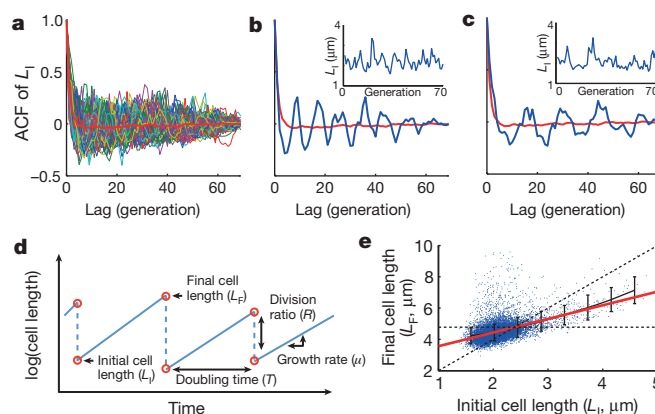


Figure 1 | Transient oscillations in cell size. **a**, ACF of all lineages ($n = 160$) and their average (thick red line). **b**, An example of L_I oscillation with a period of ~ 8 generations. **c**, Another example of L_I oscillation with a period of ~ 16 generations. **d**, Definition of growth parameters. **e**, A noisy linear map in cell size control: final cell size (L_F) is plotted against initial cell size (L_I) ($n = 11,168$). Black and red lines show binned average and linear regression line, respectively. The slope of the regression line (a) is 0.871 with a 95% confidence interval of 0.842–0.901. The two dotted lines show $y = 2x$ and $y = 2b/(2 - a)$. The error bars indicate the standard deviation of each bin.

¹Department of Biomedical Engineering, Duke University, Durham, North Carolina 27708, USA. ²Department of Physics, Duke University, Durham, North Carolina 27708, USA. ³Department of Biology, Duke University, Durham, North Carolina 27708, USA. ⁴Computational Biology and Bioinformatics, Duke University, Durham, North Carolina 27708, USA. ⁵Center for Genomic and Computational Biology, Duke University, Durham, North Carolina 27708, USA.

*These authors contributed equally to this work.

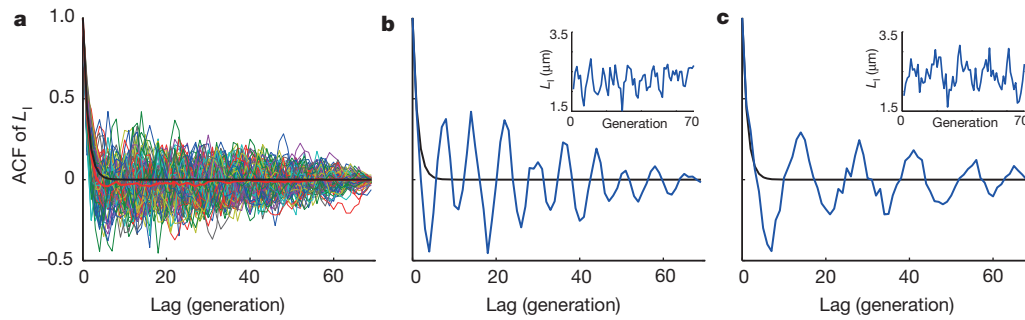


Figure 2 | Simulated transient cell-size oscillations using the noisy linear map (equation (1)). **a**, ACF of L_1 for 100 simulations (thin lines). The thick red line shows the average ACF; the thick black line shows the theoretical calculation. $a = 0.871$, $b = 2.70 \mu\text{m}$, and σ_1 (standard deviation of η) = 0.548

each for 70 generations (that is, typical duration in our experiment). The ensemble average ACF of the simulated L_1 dynamics followed a simple exponential decay, consistent with the experimental observation and theory⁷, $R(\tau) = (a/2)^{|\tau|}$ (Fig. 2a). However, some of the individual realizations showed distinct oscillations with variable periods (Fig. 2b, c). A time-frequency analysis of the simulated L_1 dynamics over longer duration (Methods) showed that the dominant frequency in a single lineage changed over time (Extended Data Fig. 3a). This indicates that the observed oscillations were transient and could arise and disappear in single lineages.

How cell-size control affects the frequency and amplitude of these apparent oscillations is unclear. As indicated by a rescaled equation (Methods, equation (2)), the noisy linear map has a single free parameter, a , that dictates the overall temporal dynamics. Biologically, a represents the strength of cell-size control; its value can be experimentally measured and it is probably determined by the molecular mechanisms underlying cell-size control^{8–11} (Supplementary Information) and growth conditions. For $a = 0$, the cell size in one generation is not influenced by that in the previous generation (that is, very strong regulation); thus the cell-size dynamics are determined by the noise term. For $0 < a < 2$, the cell size in one generation retains a ‘memory’ of the previous generation (that is, weaker regulation). a cannot exceed 2 because otherwise cell size will grow or shrink without bound. To investigate the effect of a on the transient oscillations, we simulated the cell-size dynamics for 70 generations using the rescaled linear map (equation (2)) with different a values, and compared the probability of transient oscillations and frequency of individual cells (Methods).

The simulation showed that the probability of transient oscillations was negligible when a was close to 0; the dynamics were dictated by the noise term (Fig. 3a). As a increased, the probability of transient oscillation also increased and peaked at around $a = 1.3$, above which it started to decline. The linear map acts as a low-pass filter to η (Methods and equation (3))⁷. When $a = 0$, there is no filtering and the system contains all frequencies, on average, at the same power. As a increases, the system suppresses high-frequency components and concentrates the power to the low-frequency domain, and thus slowing down the dynamics (Extended Data Fig. 3b). This filtering can generate transient low-frequency oscillations in some cells. The subsequent decline can be explained by considering how a affects the timescale of the dynamics. Our finite observation window (70 generations) limits the lowest observable oscillation frequencies, and so an extreme slow-down of the dynamics reduces the probability of oscillation for large a . The effect of this slowing down is also seen in the oscillation frequency, where the oscillation frequency decreases with a and levels off owing to the limited observation time window (Fig. 3b and Extended Data Fig. 3c).

The noisy linear map provides a simple explanation for the experimentally observed oscillations: both the probability of oscillation and the range of oscillation frequencies analysed from our experiments show excellent agreement with the simulation (Fig. 3a, b, red symbols).

μm were used for these simulations. These values are based on the characterization in Fig. 1e and the root-mean-square error of the regression was used for σ_1 . **b**, **c**, ACF of two typical oscillatory time courses (insets).

We observed similar oscillations in previously published data sets in different strains³, and they are also consistent with our simulation, although to a lesser degree (Fig. 3a, b, blue symbols). The agreement between simulation and experiment was improved when aberrant cell growth such as spontaneous cell filamentation was excluded from the data set (Extended Data Fig. 4). Finally, a close look at the distribution of oscillation frequencies reveals the dominance of low frequencies for both simulations and experimental data (Fig. 3c). Taken together, our analyses support the notion that the noisy linear map can explain the spontaneous transient oscillations in cell size.

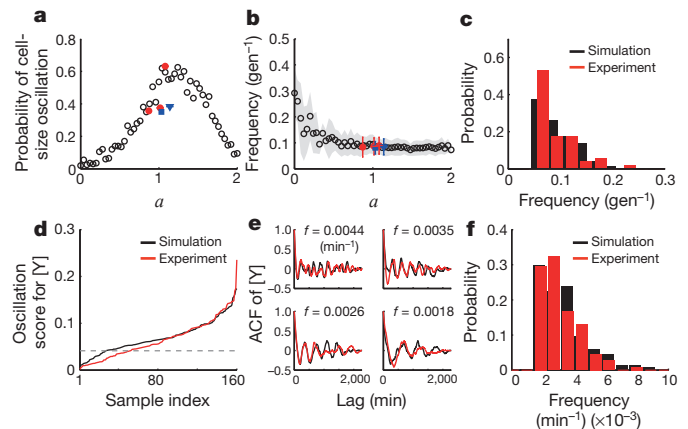


Figure 3 | Dependence of cell size oscillation on a , and oscillations in gene expression. **a**, Dependence of the probability of oscillation on a . For each value of a , 200 simulations were run using the rescaled linear map (equation (2)). Filled symbols indicate analysis of our own experimental data (MC4100 grown at 37 °C, 27 °C and 25 °C, red circles from left to right ($n = 143$, 48 and 57, respectively)) and previously published data (MG1655, blue square ($n = 97$); B/r strain, blue triangle ($n = 60$)). Only lineages without aberrant cell cycles are shown (Methods). The horizontal error bars indicate the 95% confidence interval. **b**, Dependence of the average oscillation frequency on a using the same data set presented in **a**. Only lineages that were considered oscillatory were used for the calculation ($n = 51$ (37 °C), 18 (27 °C), and 36 (25 °C) for MC4100, 34 for MG1655, and 23 for B/r strain). The shaded region represents standard deviation. As in **a**, closed symbols indicate analysis of experimental data; the vertical error bars indicate standard deviation. **c**, Distributions of oscillation frequencies from experimental (MC4100, 37 °C; $n = 51$) and simulated ($n = 80$) cell-size dynamics. Simulations were done with $a = 0.88$ using equation (2). **d**, Oscillation scores of simulated (black, Methods) and experimental (MC4100 at 37 °C, red) YFP concentrations ([Y]). The data were sorted according to the oscillation score and shown in an ascending order. The dashed line indicates a threshold for the oscillation score (Methods). **e**, ACFs of [Y] for four different frequencies. For each frequency (indicated in each panel), the sample with the highest score (above the threshold) is shown for simulation (black) and experiment (red). **f**, Distributions of oscillation frequencies in [Y] for simulation (black, $n = 134$) and experiment (red, $n = 108$).

The same principle of noisy linear map is also applicable to molecules in the cell. Consider constitutive production of a protein with negligible degradation. At steady-state growth, the quantity of the protein on average doubles during a cell cycle and halves after cell division, generating dynamics similar to cell size. Dilution of molecules by cell division inherently acts to maintain a steady-state level of the molecule. In our experiment, the yellow fluorescent protein (YFP) was constitutively expressed in cells. Indeed, analysis of the experimental data revealed a noisy linear map similar to that for cell size (that is, $0 < a = 1.05 < 2$; Extended Data Fig. 5). As such, tracking the amount of the total per-cell protein at the beginning of each cell cycle should also reveal transient oscillations, just as observed in L_i . However, it is not evident whether the protein concentration ($[Y]$) would also oscillate. To examine this aspect, we simulated the stochastic gene expression in each cell coupled with long-term cell-size dynamics using a noisy linear map (Methods). Our simulations indeed predicted transient oscillations in the protein concentration (Fig. 3d, e, black lines).

Consistent with the model prediction, analysis of experimental data revealed transient oscillations in $[Y]$ (Fig. 3d, e, red lines). The periods of these oscillations spanned many cell cycles and could be close to 20 cell cycles (Fig. 3e, f; period of ~ 600 min where average doubling time is ~ 33 min). We confirmed this observation in other growth conditions (Extended Data Figs 6 and 7) and previously published data sets³ that used other *E. coli* strains under different experimental conditions (Extended Data Figs 8 and 9). Despite the different experimental settings, our analysis consistently revealed oscillations in cell size and gene expression in each data set. This consistency indicates the generality of our conclusions.

We further probed the generality of the noisy linear map in other data sets that examined (1) *E. coli* growth in another type of microfluidic device under three different growth media¹², (2) growth of another rod-shaped bacterium, *Bacillus subtilis*¹², and (3) growth of a rod-shaped fission yeast. All these data sets revealed the existence of a noisy linear map (Fig. 4 and Extended Data Fig. 10). A recent study

also showed a consistent result for the fission yeast¹³. We note that the cell-size control strength, a , is generally not equal to 1; instead, it varies depending on growth conditions, strains, and species (Fig. 4). This conclusion differs from the ‘adder’ model^{5,6,14}, which states that cells add a constant volume (or mass) between divisions. The adder model requires $a = 1$. Our analysis suggests that the adder model^{5,6} may represent a special case of cell-size control and might not be generally applicable to different bacterial strains, species or growth conditions.

Our work reveals a simple model of cell-size control and its physiological consequence in gene expression. The strength of cell-size control may vary according to underlying molecular mechanisms and growth conditions (Fig. 4 and Supplementary Information). This change would lead to varying degree of spontaneously generated pulsatile gene expression, which has been implicated in stress response, signalling, and development^{15–17}. As cell-size control is a fundamental aspect of biology, it may represent a primitive means to generate spontaneous pulsatile gene expression for these functions. Also, recent studies^{18–20} found that the bacterial proteome is partitioned into distinct sectors, which can be adjusted to optimize resource allocation under different growth conditions. Our observation of spontaneous, long-term oscillations in gene expression suggests that such balance may be adjusted dynamically even under a fixed growth condition. As the bacterial ‘growth law’ revealed by the previous work was based on population-level measurements^{18–20}, it would be interesting to investigate this empirical law at the single-cell level. Efforts have already been made to determine how fluctuation in catabolic gene expression affects cell growth and vice versa²¹. Finally, the noisy linear map could be exploited in interfacing with synthetic gene circuits. Studies have shown that interactions between host physiology and exogenous gene circuits can lead to new behaviours^{22–24}. Given the ability of the noisy linear map to generate transient oscillations in gene expression, it would be interesting to examine its effects on synthetic oscillators^{25–27}.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 April 2014; accepted 14 May 2015.

Published online 3 June 2015.

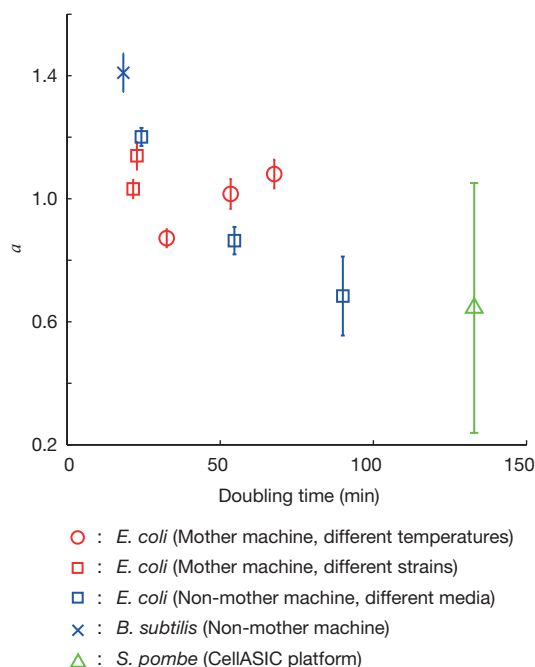


Figure 4 | a in different data sets. The slope (a) of the linear map was derived from 10 different data sets and plotted as a function of cell doubling time. ‘Mother machine’ refers to our own data sets (‘different temperatures’ and the data sets from ref. 3 (‘different strains’). ‘Non-mother machine’ refers to the data set from ref. 12. ‘CellASIC platform’ refers to the *S. pombe* experiment performed using the CellASIC system (a commercially available microfluidic device). The error bars indicate 95% confidence intervals.

- Raj, A. & van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135**, 216–226 (2008).
- Huh, D. & Paulsson, J. Random partitioning of molecules at cell division. *Proc. Natl Acad. Sci. USA* **108**, 15004–15009 (2011).
- Wang, P. *et al.* Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103 (2010).
- Osella, M., Nugent, E. & Cosentino Lagomarsino, M. Concerted control of *Escherichia coli* cell division. *Proc. Natl Acad. Sci. USA* **111**, 3431–3435 (2014).
- Taheri-Araghi, S. *et al.* Cell-size control and homeostasis in bacteria. *Curr. Biol.* **25**, 385–391 (2015).
- Campos, M. *et al.* A constant size extension drives bacterial cell size homeostasis. *Cell* **159**, 1433–1446 (2014).
- Kitagawa, G. *Introduction to Time Series Modeling* 31–48 (Chapman and Hall/CRC, 2010).
- Donachie, W. D. & Blakely, G. W. Coupling the initiation of chromosome replication to cell size in *Escherichia coli*. *Curr. Opin. Microbiol.* **6**, 146–150 (2003).
- Reyes-Lamothe, R., Nicolas, E. & Sherratt, D. J. Chromosome replication and segregation in bacteria. *Annu. Rev. Genet.* **46**, 121–143 (2012).
- Erickson, H. P., Anderson, D. E. & Osawa, M. FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one. *Microbiol. Mol. Biol. Rev.* **74**, 504–528 (2010).
- de Boer, P. A. J. Advances in understanding *E. coli* cell fission. *Curr. Opin. Microbiol.* **13**, 730–737 (2010).
- Moffitt, J. R., Lee, J. B. & Cluzel, P. The single-cell chemostat: an agarose-based, microfluidic device for high-throughput, single-cell studies of bacteria and bacterial communities. *Lab Chip* **12**, 1487–1494 (2012).
- Nobs, J.-B. & Maerkl, S. J. Long-term single cell analysis of *S. pombe* on a microfluidic microchemostat array. *PLoS ONE* **9**, e93466 (2014).
- Amir, A. Cell size regulation in bacteria. *Phys. Rev. Lett.* **112**, 208102 (2014).
- Levine, J. H., Lin, Y. & Elowitz, M. B. Functional roles of pulsing in genetic circuits. *Science* **342**, 1193–1200 (2013).
- Süel, G. M., Kulkarni, R. P., Dworkin, J., Garcia-Ojalvo, J. & Elowitz, M. B. Tunability and noise dependence in differentiation dynamics. *Science* **315**, 1716–1719 (2007).
- Wakamoto, Y. *et al.* Dynamic persistence of antibiotic-stressed mycobacteria. *Science* **339**, 91–95 (2013).

18. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1102 (2010).
19. You, C. *et al.* Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature* **500**, 301–306 (2013).
20. Hui, S. *et al.* Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, 784 (2015).
21. Kiviet, D. J. *et al.* Stochasticity of metabolism and growth at the single-cell level. *Nature* **514**, 376–379 (2014).
22. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nature Chem. Biol.* **5**, 842–848 (2009).
23. Marguet, P., Tanouchi, Y., Spitz, E., Smith, C. & You, L. Oscillations by minimal bacterial suicide circuits reveal hidden facets of host-circuit physiology. *PLoS ONE* **5**, e11909 (2010).
24. Cookson, N. A. *et al.* Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Mol. Syst. Biol.* **7**, 561 (2011).
25. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
26. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519 (2008).
27. Mondragón-Palomino, O., Danino, T., Selimkhanov, J., Tsimring, L. & Hasty, J. Entrainment of a population of synthetic genetic oscillators. *Science* **333**, 1315–1319 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Jun for providing the original mother machine and P. Cluzel for providing their data set on *E. coli* and *B. subtilis* growth. We also thank the Light Microscopy Core Facility at Duke University for their help in conducting microscopy experiments. This work was partially supported by a National Science Foundation Career Award (L.Y.), the National Institutes of Health (NIH) (L.Y., R01GM098642, R01GM110494), a DuPont Young Professorship (L.Y.), a David and Lucile Packard Fellowship (L.Y.), a DARPA Biochronicity Grant (DARPA-BAA-11-66, N.E.B.), a NIH Director's New Innovator Award (DP2 OD008654-01, N.E.B.), and a Burroughs Wellcome Fund CASI Award (BWF 1005769.01, N.E.B.).

Author Contributions Y.T. conceived the research, designed and performed both modelling and experimental analyses, interpreted the results, and wrote the manuscript. A.P. conceived the research, designed and performed experimental analyses, interpreted the results, and wrote the manuscript. H.P. designed and performed experimental analyses and interpreted the results. S.H. fabricated the microfluidic device and performed experiments. R.S. developed the software for image analysis. N.E.B. assisted in data interpretation and manuscript revisions. L.Y. conceived the research, assisted in research design and data interpretation, and wrote the manuscript. All authors approved the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.Y. (you@duke.edu).

METHODS

Fabrication of microfluidic device. We followed the previously published procedure to fabricate the ‘mother machine’³ except that our mould was reverse-fabricated from the original mother machine device. This was done by pouring epoxy onto the original mother machine device (gift from S. Jun). Replicas of the mother machine were then created by pouring PDMS onto this mould and solidifying the polymer at 80 °C for 30 min. The resulting PDMS device was then bonded to a glass cover slip by plasma treatment.

Cell strain, growth condition, microscopy and microfluidics. An *E. coli* strain MC4100 that constitutively expresses YFP (*galK::P_{lac}-yfp amp^R*; gift from R. Kishony²⁸) was used in our experiments. For long-term imaging of these cells in the microfluidic device, a similar procedure as described in ref. 3 was followed. In brief, cultures were grown overnight in LB at 37 °C. The overnight culture was diluted 100-fold in 5 ml fresh LB and grown at 37 °C. At sufficiently high density, this culture was concentrated about 20-fold by centrifugation for loading into the mother machine using a syringe. The device was then spun for 3 min using a mini centrifuge to help trap cells in the side channels of the device. Before loading, the device was cleaned using 70% ethanol, washed twice with distilled water, and all liquid was then expunged with air. Fresh LB was then introduced to remove cells not trapped and a continuous flow (100 µl h⁻¹) was maintained. Throughout the experiment, carbenicillin (50 µg ml⁻¹) was added in growth medium. Images were acquired at one-minute intervals using DeltaVision Elite microscope (Applied Precision) with a motorized stage and an Evolve EM-CCD camera (Photometrics) with either ×100 DIC objective or ×60 phase objective. When ×60 phase objective was used, additional ×2 auxiliary magnification was also used. Before the experiment, the microscope and its growth chamber were equilibrated at an appropriate temperature (25, 27 or 37 °C), and the temperature was maintained throughout the experiment.

For the experiment with *S. pombe*, JM1645 strain that constitutively expresses GFP (*pAct1-GFP::leu1-32 h-*; gift from J. Moseley) was grown overnight in 3 ml YE4S medium at 30 °C. The overnight culture was diluted 100-fold into fresh 3 ml YE4S and grown at 30 °C for 4.5 h. Cells were then sonicated for 30 s to separate at medium intensity (Diagenode Bioruptor UCD-200). Cells were loaded into a CellASIC Y04C plate (EMD Millipore). A medium flow rate was kept at 3 p.s.i., and the middle chamber (4 µm in height) was used for imaging. Images were acquired every 3 min using DeltaVision Elite microscope with a motorized stage and a CoolSNAP CCD camera (Photometrics) with ×60 phase objective. The temperature was kept at 30 °C throughout the experiment.

Image analysis. For mother machine experiments, we developed a custom program using C++ and FIJI for image segmentation. The segmentation of cells was performed based on fluorescent images by finding minima of fluorescent intensity along the channel direction of the mother machine. The segmented images were checked manually and corrected. CellStat (Fraunhofer-Chalmers Centre)²⁹ was used for image segmentation of the experiment with *S. pombe*.

Data processing (mother machine data). For both our own data sets and previously published data sets, cell divisions were detected based on the change in cell length—a division event is identifiable in the data sets as a clear and large drop in the size of the mother cell. We selected cell lineages that contained measurements of full 70 generations. This resulted in 160 (37 °C), 54 (27 °C) and 65 (25 °C) lineages for MC4100, 158 lineages for MG1655, and 80 lineages for B/r strain. To minimize the effect of erroneous segmentation, we ignored spontaneous ‘spikes’ in cell length, which were occasionally observed in the previously published data³. Cells occasionally undergo aberrant cell growth such as filamentation. For the analysis performed in Fig. 1e and Extended Data Figs 1, 2 and 5, we excluded these instances, by discarding data points in which (1) initial cell length was larger than $\bar{L} + 2\sigma_L$ (\bar{L} and σ_L are the average and standard deviation of the cell size distribution, respectively); or (2) final cell length was larger than $2(\bar{L} + 2\sigma_L)$. We also excluded cell cycles whose initial cell length was smaller than $\bar{L} - 2\sigma_L$. This latter condition was to mainly exclude data points that appear to result from erroneous segmentation in the previously published data sets. Our own data sets (that is, MC4100) contained only one such instance. To eliminate incorrect segmentations, we assumed cell divisions to be at least 10 min apart and excluded cell cycles in which $L_F < L_I$ (only found in the published data sets). These procedures filtered out ~0.6% of total cell cycles. For the analysis performed in Fig. 3a–c and Extended Data Fig. 4, unless otherwise noted, cell lineages that contained aberrant cell cycles described above were excluded from the analysis.

We note that the published data sets were available and accessed in the form of processed data sets rather than the actual movies of cell growth themselves. As such, we were unable to correct manually for possible segmentation errors other than using the criteria described above. This inability could account for lower probability of oscillation as compared with the model prediction (Fig. 3a); the cell size dynamic is more susceptible to these errors than the linear map. Indeed, when the above criteria were used to process the data, it led to a clear improvement in

agreement with the simulations (Extended Data Fig. 4). Our own data sets were subject to manual segmentation checks based on the movies and were largely devoid of these errors.

Data processing (data from ref. 12). For each condition (*E. coli* with three growth media and *B. subtilis*), we chose to analyse an experiment performed with 4% agarose. All complete cell cycles were subject to the above processing, and the same exclusion criteria as the mother machine data were uniformly applied for filtering. As a result, ~0.6% of total cell cycles was filtered out.

Data processing (*S. pombe* experiment). For the *S. pombe* experiment, microscope images were analysed using CellStat (Fraunhofer–Chalmers Centre)²⁹. Cell area, instead of cell length, was used to derive the linear map. Since this experiment was performed in the CellASIC platform and the throughput was much lower than the mother machine (87 cell cycles), apparent filamentous events were manually inspected and two cell cycles were excluded. These two events were visually obvious and well separated from the other cell cycles in the distribution. The total of 85 cell cycles were used for the linear map analysis (Extended Data Fig. 10e).

Analysis of noisy linear map. The noisy linear map (equation (1) in the main text)

$$L_I(n+1) = (aL_I(n) + b + \eta)/2$$

can be rescaled by

$$x(n) = \frac{L_I(n) - L_I^{ss}}{\sigma_1}$$

and

$$L_I^{ss} = \frac{b}{2-a}$$

Here, σ_1 is the standard deviation of the noise term η , and L_I^{ss} is the steady state of $L_I(n)$. This gives

$$x(n+1) = (ax(n) + \eta')/2 \quad (2)$$

where $\eta' = \eta/\sigma_1$. We assume a Gaussian white noise for η , and then η' is also a Gaussian white noise with $\eta' \sim N(0, 1)$. This rescaling shows that the dynamics of noisy linear map is solely dependent on a . For the analysis performed in Fig. 3a–c, we used this rescaled model to examine the effect of a on oscillation characteristics.

As discussed in the main text, the autocorrelation function of the system is given by $R(\tau) = (a/2)^{|\tau|}$ and then its power spectrum, $X(f)$, for $0 \leq a \leq 2$ is given by⁷

$$X(f) = |H(f)|^2 S_{\eta}(f) = \frac{1}{1 - a \cos 2\pi f + (\frac{a}{2})^2} \quad (3)$$

in which $H(f)$ is a transfer function of the system (at frequency, f), and $S_{\eta}(f) = 1$ is the power spectrum of η' . With $a = 2$, the dynamics becomes Brownian noise, and $X(f) = 1/(2\pi f)^2$. As a increases, the system suppresses high frequency components and concentrates the power to the low frequency domain (Extended Data Fig. 3b).

Time-frequency analysis of L_I dynamics. To gain further insights into L_I dynamics, we performed a time-frequency analysis. Specifically, we simulated the process for a much longer duration (700 generations as opposed to the 70 generations that we could observe experimentally) using equation (1), and then calculated power spectrum density (PSD) for each period of 70 generations with 50% overlap. That is, PSDs were calculated for data from 1 to 70 generations, 36 to 105 generations, 71 to 140 generations, and so on. A single frequency occasionally dominated the system and then disappeared, creating a patchy appearance of the plot (Extended Data Fig. 3a). Consistent with our experimental observations (Fig. 1b, c) across an ensemble of single cells, the dominant frequency in a single cell changed across time. In some time windows, no dominant frequency was observed, which indicates that the observed oscillations are transient and can arise and disappear in single cells or across an ensemble of cells (Extended Data Fig. 3a).

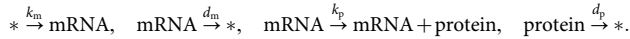
Simulation of gene expression under cell size regulation. To examine gene expression dynamics in the presence of cell size regulation by simulation, we combined the noisy linear map and stochastic gene expression model. For each simulation, we first generated cell size dynamics based on the noisy linear map as follows:

1. Compute $L_F(n)$ using $L_F(n) = aL_I(n) + b + \eta$ but ensure $L_F(n) > L_I(n)$
2. Assuming a constant growth rate, μ , construct cell size profile using $L(t) = L_I(n)e^{\mu t}$, $0 \leq t \leq T$ where $L_I(n)e^{\mu T} = L_F(n)$.
3. Compute $L_I(n+1)$ using $L_I(n+1) = L_F(n)R$ where $R \sim N(0.5, \sigma_2)$, but ensure $0 < R < 1$.
4. Iterate steps 1–3.

Two bounds were applied to the noisy linear map model to accommodate biological constraints. First, in step 1 we ensured that the final cell size $L_F(n)$ is larger than the initial cell size $L_I(n)$, and that $T \geq 10$ min. Second, in step 3 the division was made stochastic within the boundary $0 < R < 1$. In Fig. 3d–f, the parameter values for a , b , σ_1 (standard deviation of η), and σ_2 were derived from the experimental observations

of the MC4100 strain grown at 37 °C ($a = 0.871$, $b = 2.70$, $\sigma_1 = 0.548$, and $\sigma_2 = 0.0344$).

After generating a cell size profile, the dynamics of constitutive gene expression was simulated using the Gillespie algorithm with following reactions:



In these simulations, we assumed: $k_m = 2$ molecules min^{-1} , $d_m = 0.2$ min^{-1} , $k_p = 0.1$ min^{-1} , and $d_p = 0.001$ min^{-1} . After cell division, mRNA and protein molecules were binomially distributed to the progeny based on the division ratio (R), and the simulation continued to the next cell cycle. To make the analysis of simulation results comparable to that of experimental data, the simulation results were sampled at the same time resolution as the experimental data.

Power spectrum analysis and scoring of oscillation. To quantify the goodness of oscillation and extract a main frequency of the oscillation, we develop a scoring system based on power spectrum analysis. First, a PSD estimate ($S(f_i)$) of temporal data of length l was computed for individual courses using Welch's method (pwelch function in Matlab). The number of segments and overlap used were 3 and 50–53%, respectively (for the analysis of L_1 ($l = 70$), the overlap of $19/36 \approx 53\%$ was used to have three complete segments, but for the analysis of [Y] the overlap was 50%). A maximum peak ($S(f_c)$) is then found and the score (z) is calculated as follows:

$$z = \frac{S(f_c)}{\sum_i^N S(f_i)} \left(1 - \frac{H(S)}{H_{\max}} \right),$$

$$\text{with } H(S) = - \sum_{i \text{ for } S(f_i) \leq S(f_c)}^N P(f_i) \log_2 P(f_i)$$

$$\text{and } H_{\max} = -\log_2 \frac{1}{M},$$

where N is the number of bins resulting from the discrete Fourier transform, $M \leq N$ is the number of bins where $S(f_i) \leq S(f_c)$ and $i > 1$ (that is, non-DC (non-constant) component), and $H(S)$ is the information entropy. The first and second terms of z represent the dominance and peakiness of the maximum peak, respectively, and f_c corresponds to the oscillation frequency. We note that the first non-DC component was not considered as a peak, and thus the slowest oscillation frequency f_{\min} detected by this method is the second non-DC component in the PSD (that is, $2/N$). Likewise, the fastest oscillation frequency f_{\max} corresponds to the second to last component in the PSD (that is, $1/2 - 1/N$).

To separate oscillatory dynamics from non-oscillatory ones, we needed to set a threshold for the oscillation score. To this end, we determined this threshold, z_c , based on the scores of noisy sine waves with various frequencies:

$$x_t = \sin(2\pi\phi_j t) + \varepsilon(t = 1, 2, \dots, l \text{ and } \varepsilon \sim N(0, 1)),$$

where, $f_{\min} \leq \phi_j \leq f_{\max} = \max(f_i)$ ($i = 1, 2, \dots, N$). We sampled the total of 101 different frequencies (that is, $j = 1, 2, \dots, 101$ and $\phi_{j+1} - \phi_j = (f_{\max} - f_{\min})/100$). In principle, these noisy sine waves represent true oscillatory dynamics. We ran 100 simulations for each ϕ_j , and calculated the average ($\mu_z(\phi_j)$) and standard deviation ($\sigma_z(\phi_j)$) of the scores. Then, the threshold z_c was defined as

$$z_c = \frac{1}{101} \sum_j^{101} (\mu_z(\phi_j) - \sigma_z(\phi_j)).$$

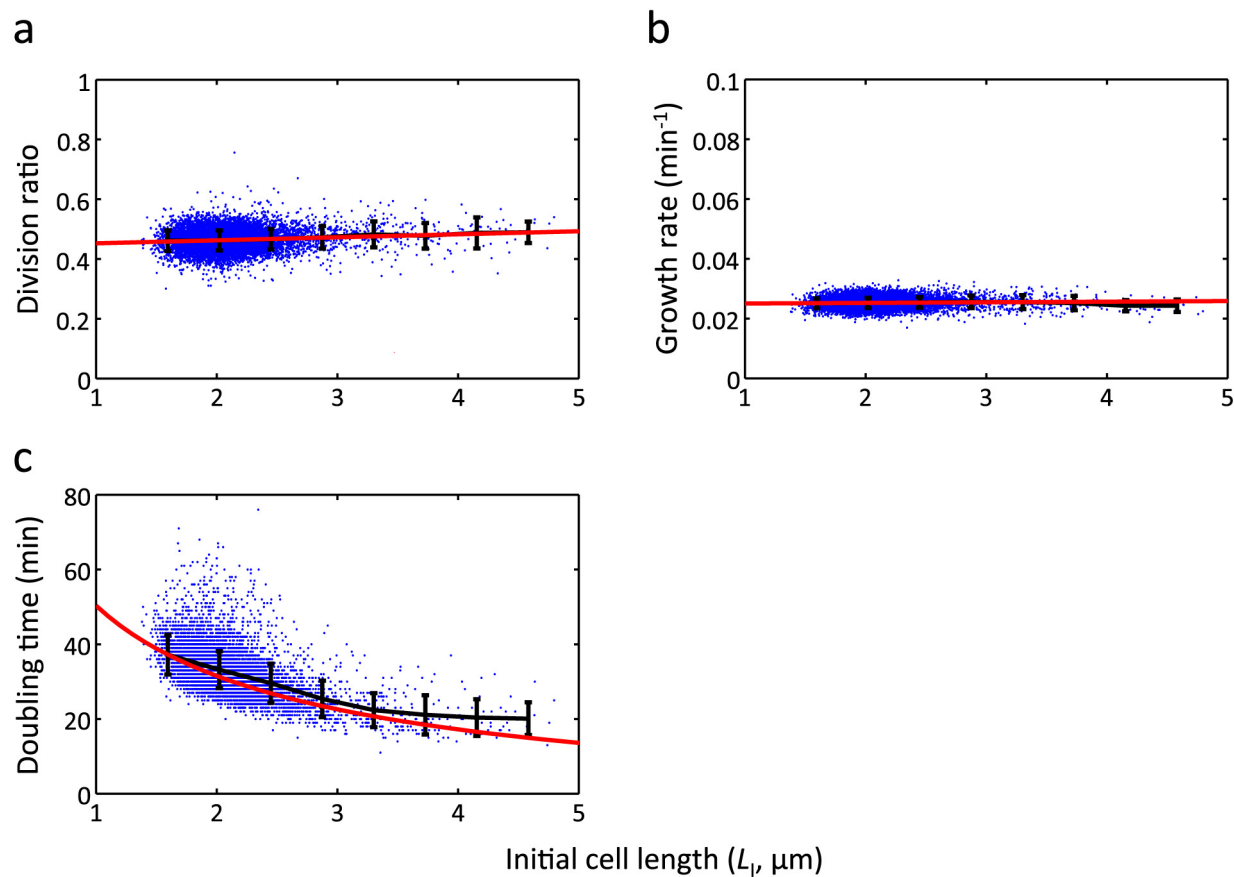
z_c varies depending on the data set (for example, different l). For L_1 dynamics, $z_c = 0.0166$, whereas for [Y] dynamics of MC4100 (37 °C), MC4100 (27 °C), MC4100 (25 °C), MG1655, and B/r, $z_c = 0.0414$, 0.0439, 0.0450, 0.0390 and 0.0394, respectively.

Dependence of cell doubling time on initial cell size. At the fundamental level, cell size control can be achieved by modulating either growth rate (μ) or doubling time (T). Our analysis shows that growth rate is relatively independent of the initial cell size (Extended Data Fig. 1a, b), but doubling time has a decreasing trend with initial cell size (Extended Data Fig. 1c). This indicates that the cell size regulation is mainly achieved via modulation of doubling time rather than growth rate. This is consistent with a recent analysis of long-term *E. coli* growth data⁴. Given the linear relationship between the initial and final cell sizes ($L_F = aL_I + b$), and assuming a constant growth rate and exponential cell growth ($L_F = L_I e^{\mu T}$) during cell cycle³⁰, we can derive an expression for the doubling time,

$$T = \ln(a + b/L_I) / \mu \quad (4)$$

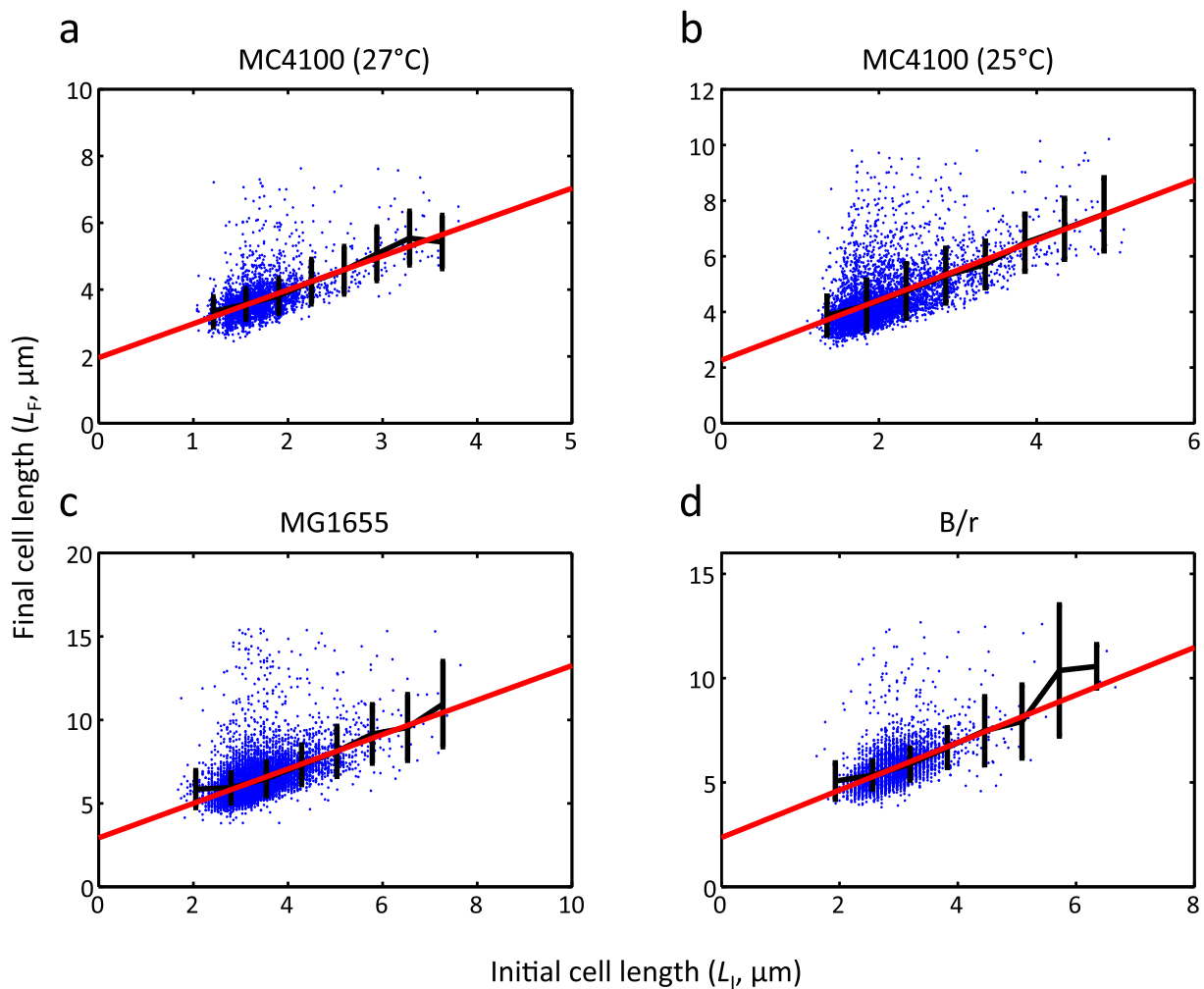
This equation shows good agreement with our experimental observation (Extended Data Fig. 1c, red line).

28. Hegreness, M., Shores, N., Hartl, D. & Kishony, R. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* **311**, 1615–1617 (2006).
29. Kvarnström, M., Logg, K., Diez, A., Bodvard, K. & Käll, M. Image analysis algorithms for cell contour recognition in budding yeast. *Opt. Express* **16**, 12943–12957 (2008).
30. Godin, M. *et al.* Using buoyant mass to measure the growth of single cells. *Nature Methods* **7**, 387–390 (2010).



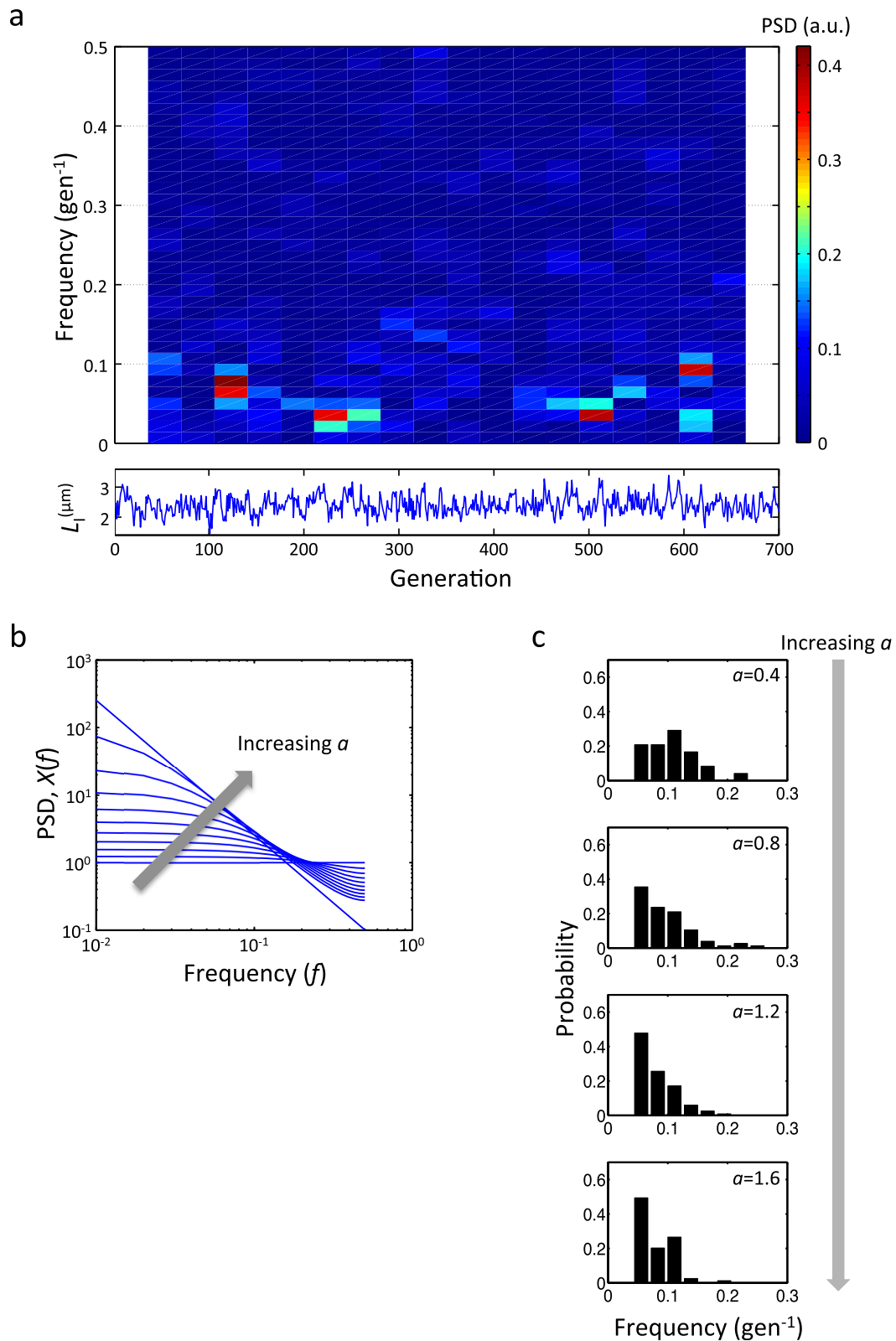
Extended Data Figure 1 | Analysis of cell growth parameters. a–c, Division ratio (a), growth rate (b) and doubling time (c) are plotted against initial cell length ($n = 11,168$). Black and red lines show binned average and trend line,

respectively. In a and b, the trend lines are the linear regression line, whereas equation (4) was used in c. The error bars indicate standard deviation of each bin.



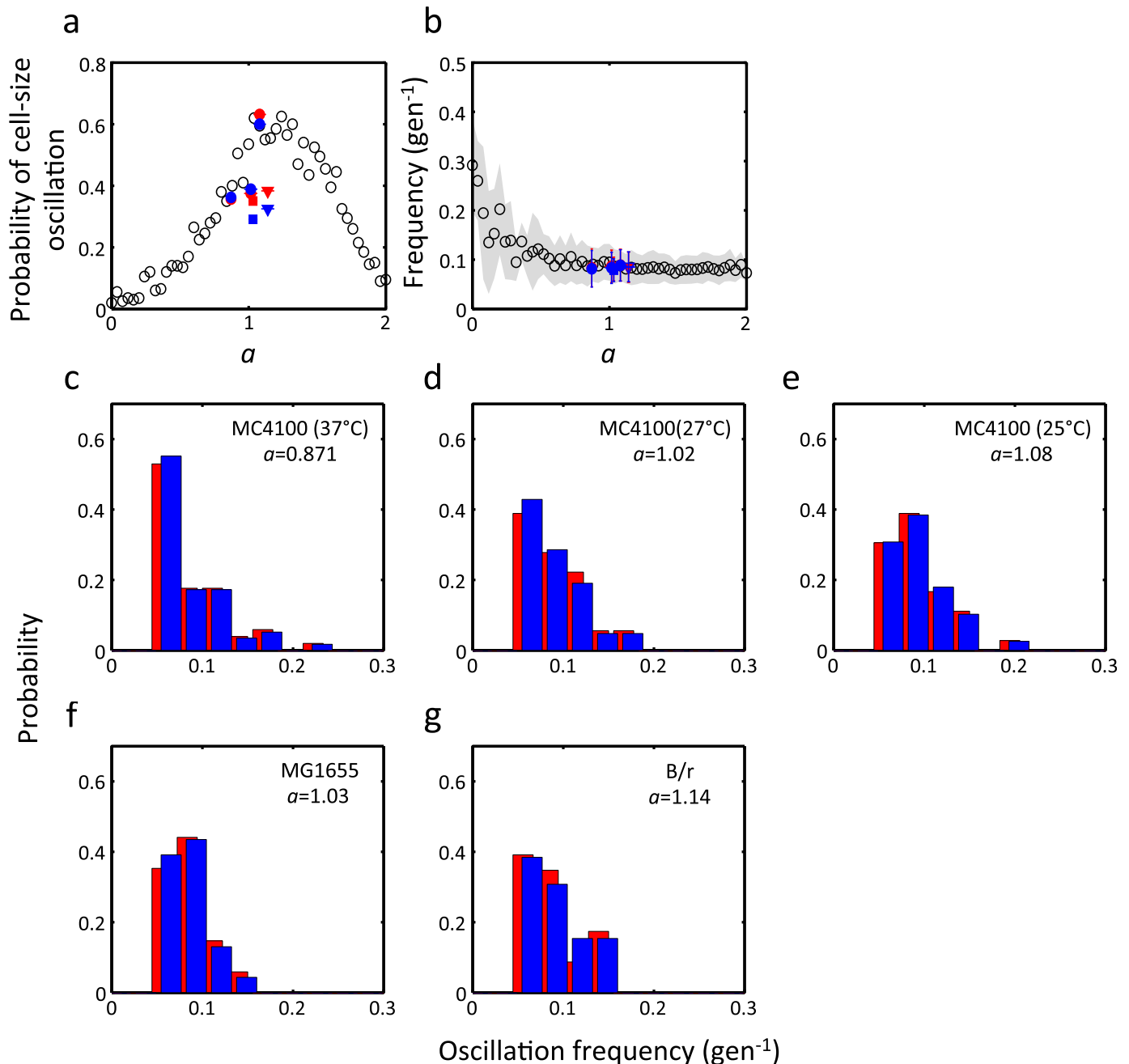
Extended Data Figure 2 | A noisy linear map in *E. coli* cell size control across different growth temperatures and strains. The same analysis as in Fig. 1e (L_I versus L_F) was performed for MC4100 grown at 27 °C (a, $n = 3,772$), MC4100 grown at 25 °C (b, $n = 4,539$), MG1655 grown at 37 °C (c, $n = 10,964$),

and B/r strain grown at 37 °C (d, $n = 5,541$). The data sets for MG1655 and B/r strain were from a previous study³. Black and red lines show binned average and linear regression line, respectively. In a–d, $a = 1.02, 1.08, 1.03$ and 1.14 , respectively. The error bars indicate standard deviation of each bin.



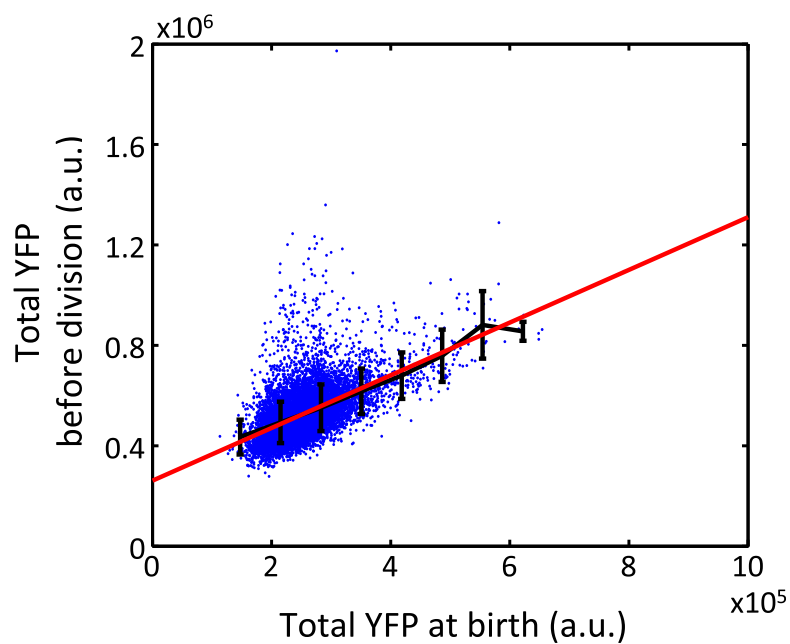
Extended Data Figure 3 | Frequency analysis of noisy linear map. **a**, Time-frequency analysis of the noisy linear map model for L_I . The top panel shows a spectrogram of 700-generation simulation constructed using 70-generation segments with 50% overlap. The bottom panel shows the temporal dynamics of L_I . The same parameters as in Fig. 2a were used. **b**, The power spectrum of noisy linear map ($X(f)$). Equation (3) is plotted for different a ($0 \leq a \leq 1.8$ with

0.2 interval). $X(f) = 1/(2\pi f)^2$ for $a = 2$ is the straight line in this log-log plot. Note that the maximum value of f is 0.5 as the time resolution is 1 generation. **c**, Dependence of oscillation frequencies on a simulated using the rescaled linear map (equation (2)). The noisy linear map model was simulated (Fig. 3a–c) and the distributions of oscillation frequencies are shown for four different values of a (from top to bottom, $n = 24, 76, 117$ and 79 , respectively).

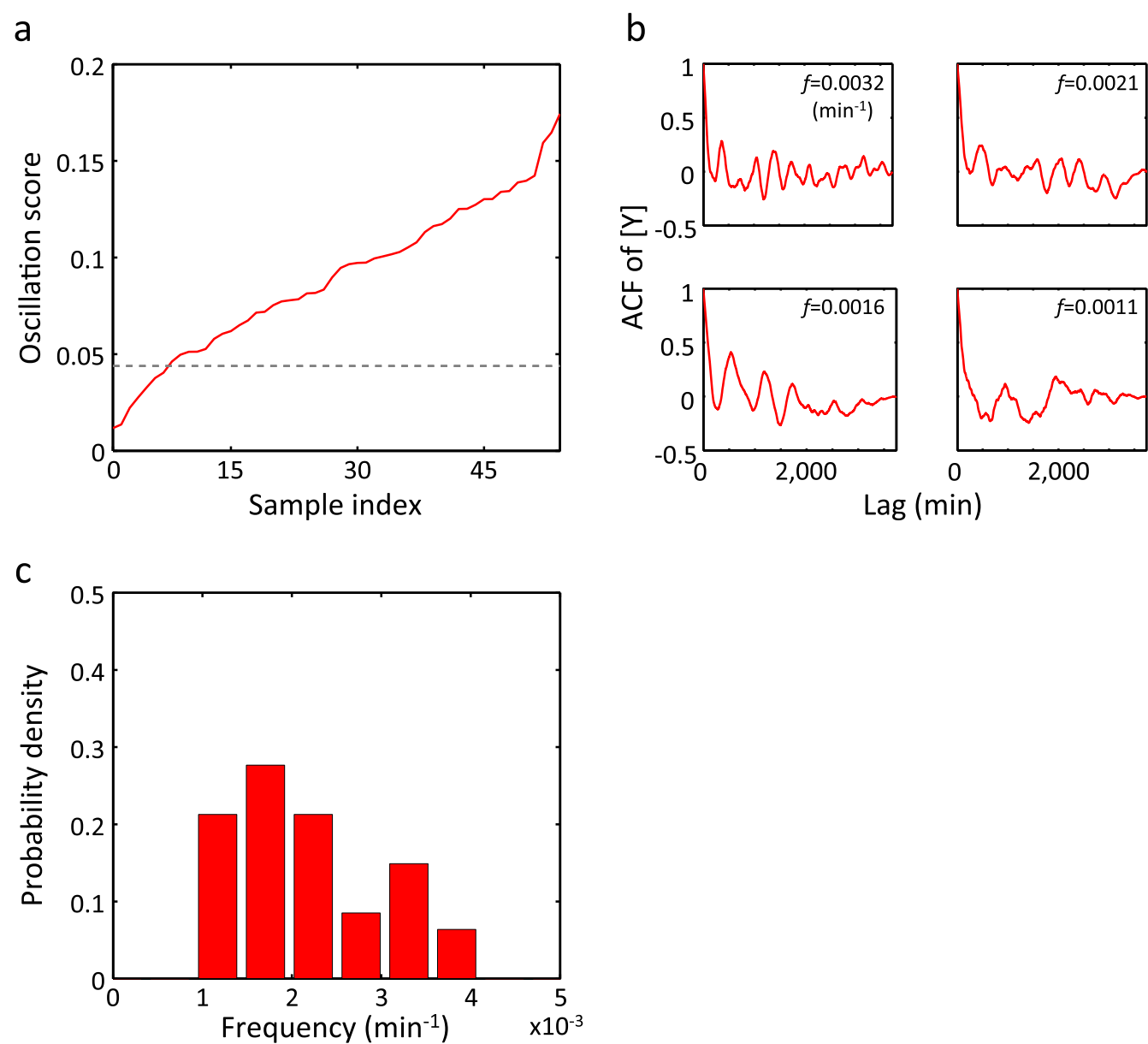


Extended Data Figure 4 | Comparison of L_1 oscillation characteristics between all lineages (blue) and lineages without aberrant cell cycles (red). a–g, Probability of oscillation (a), average oscillation frequency (b), and distributions of oscillation frequencies (c–g) are shown. In a and b, filled symbols represent experimental data (data shown in red are the same as Fig. 3a, b): circles are MC4100 grown at three different temperatures; squares and triangles are MG1655 and B/r strain, respectively. The unfilled circles were generated from simulations using the rescaled linear map (equation (2), the same plot as Fig. 3a, b). In a, the data shown in blue include 160 (37°C),

54 (27°C) and 65 (25°C) lineages for MC4100, 158 lineages for MG1655, and 80 lineages for B/r strain. The data shown in red include 143 (37°C), 48 (27°C), and 57 (25°C) lineages for MC4100, 97 lineages for MG1655, and 60 lineages for B/r strain. In b–g, only lineages that were considered oscillatory were used. For the data set in blue, $n = 58$ (37°C), 21 (27°C), and 39 (25°C) for MC4100, 46 for MG1655, and 26 for B/r strain. For the data set in red, $n = 51$ (37°C), 18 (27°C) and 36 (25°C) for MC4100, 34 for MG1655, and 23 for B/r strain.

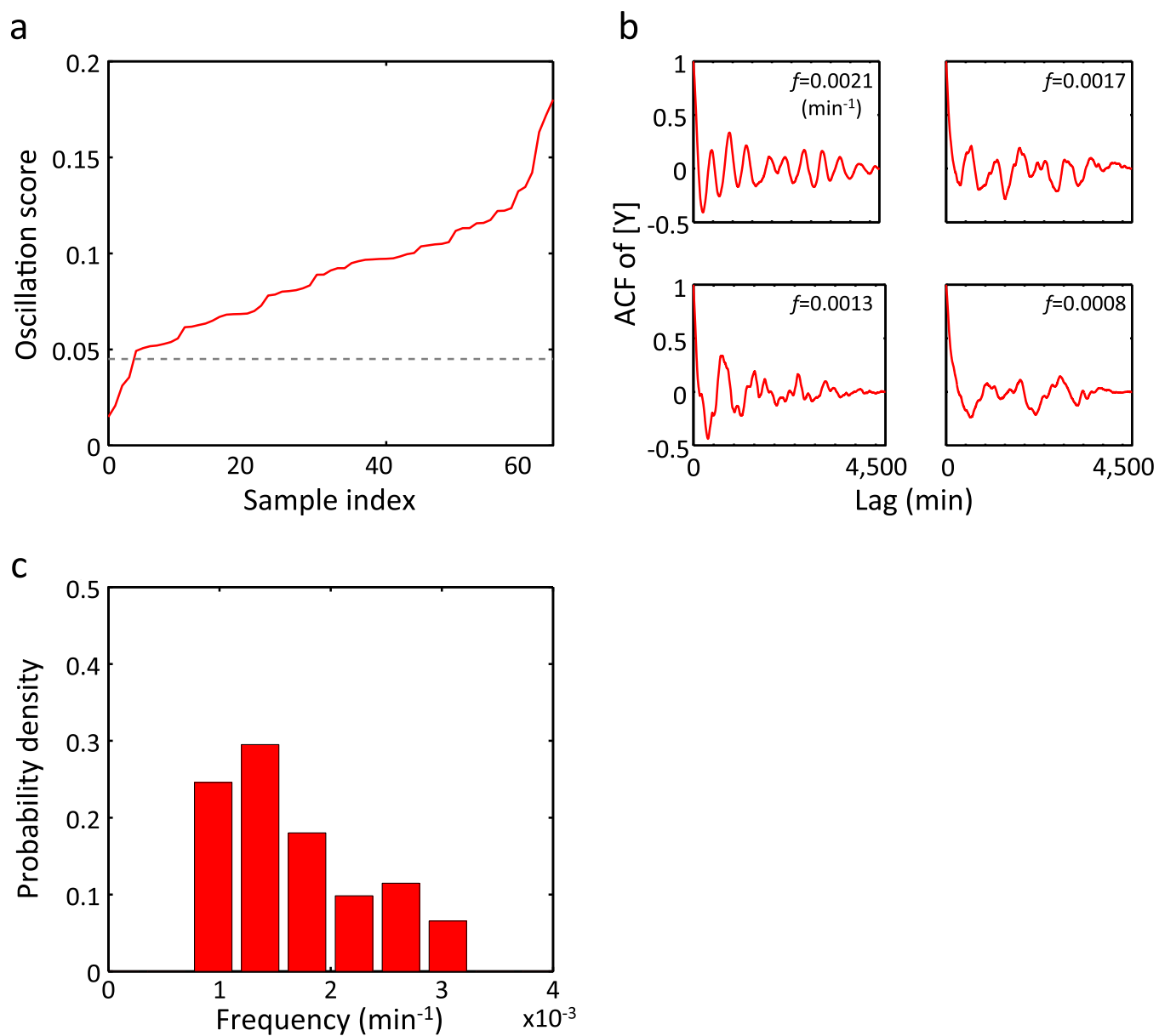


Extended Data Figure 5 | A noisy linear map in total per-cell YFP. Total YFP before division is plotted against total YFP at birth ($n = 11,168$), revealing a linear map with $a = 1.05$. Black and red lines show binned average and the linear regression line, respectively. The error bars indicate standard deviation of each bin.

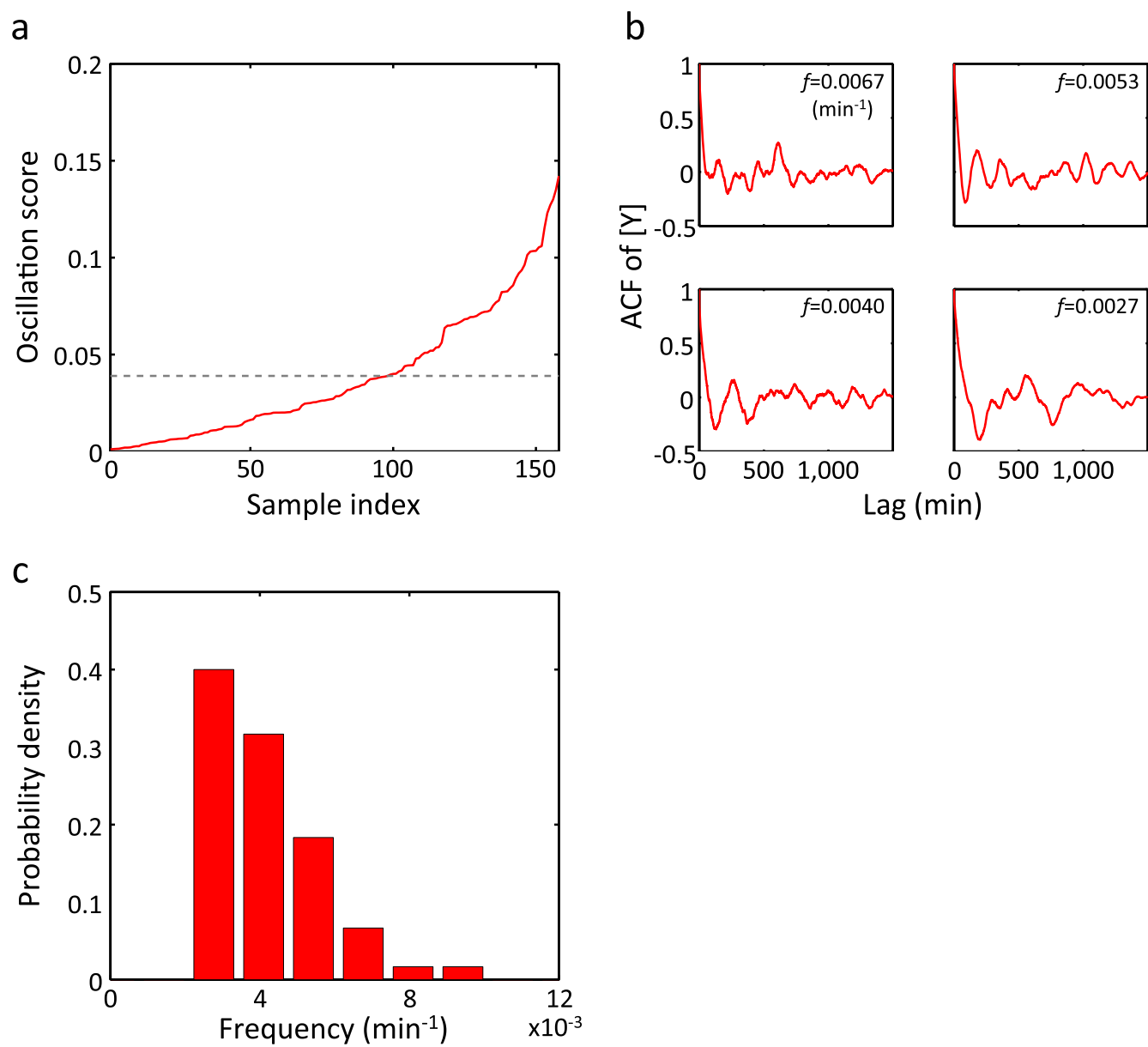


Extended Data Figure 6 | Oscillation in [Y] observed in MC4100 at 27 °C.
a, Oscillation scores of YFP concentration are shown in an ascending order. The dashed line indicates a threshold for oscillation. **b**, For four different

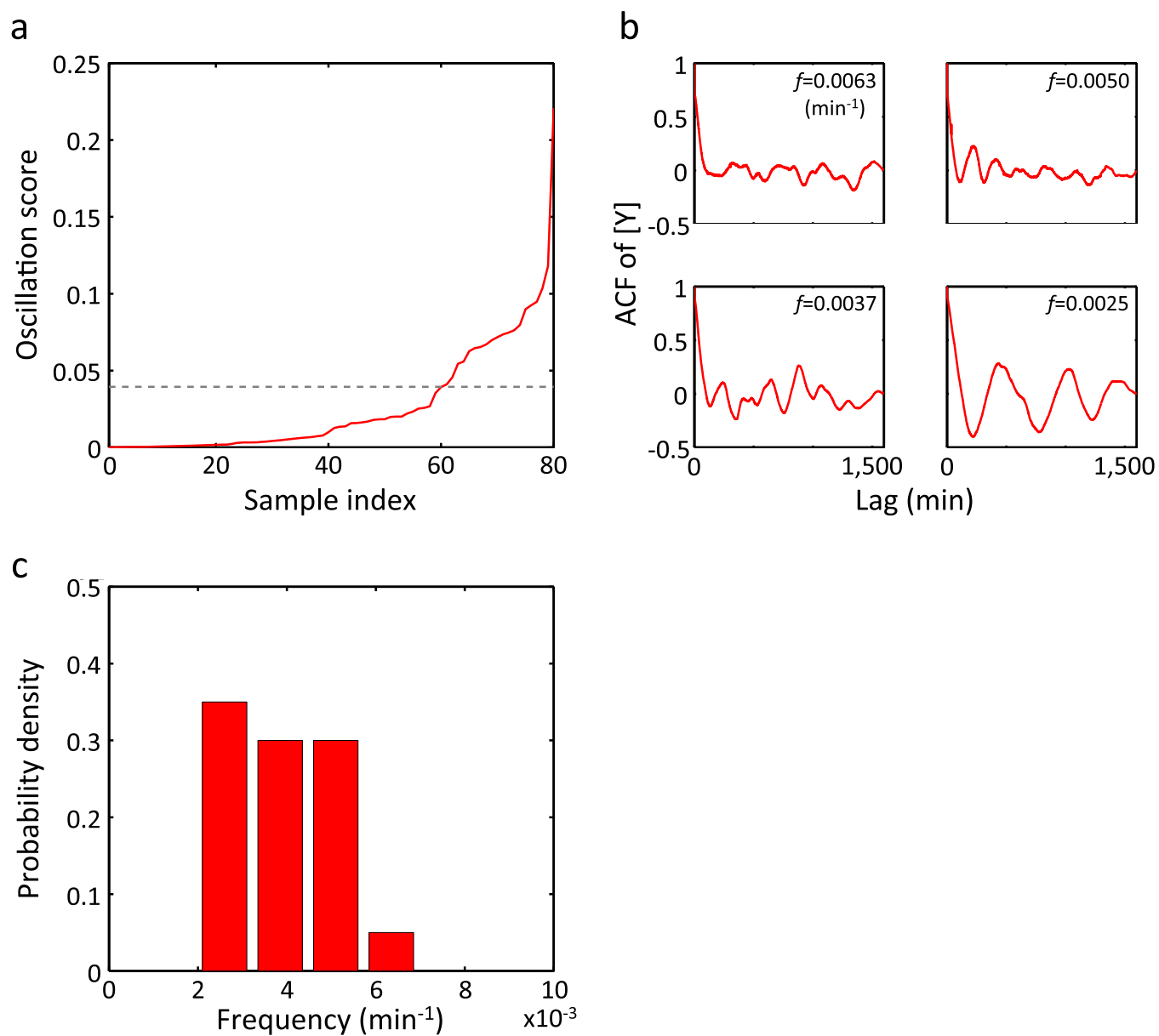
oscillation frequencies, ACFs with the highest oscillation scores are shown. **c**, The distribution of oscillation frequencies ($n = 47$).



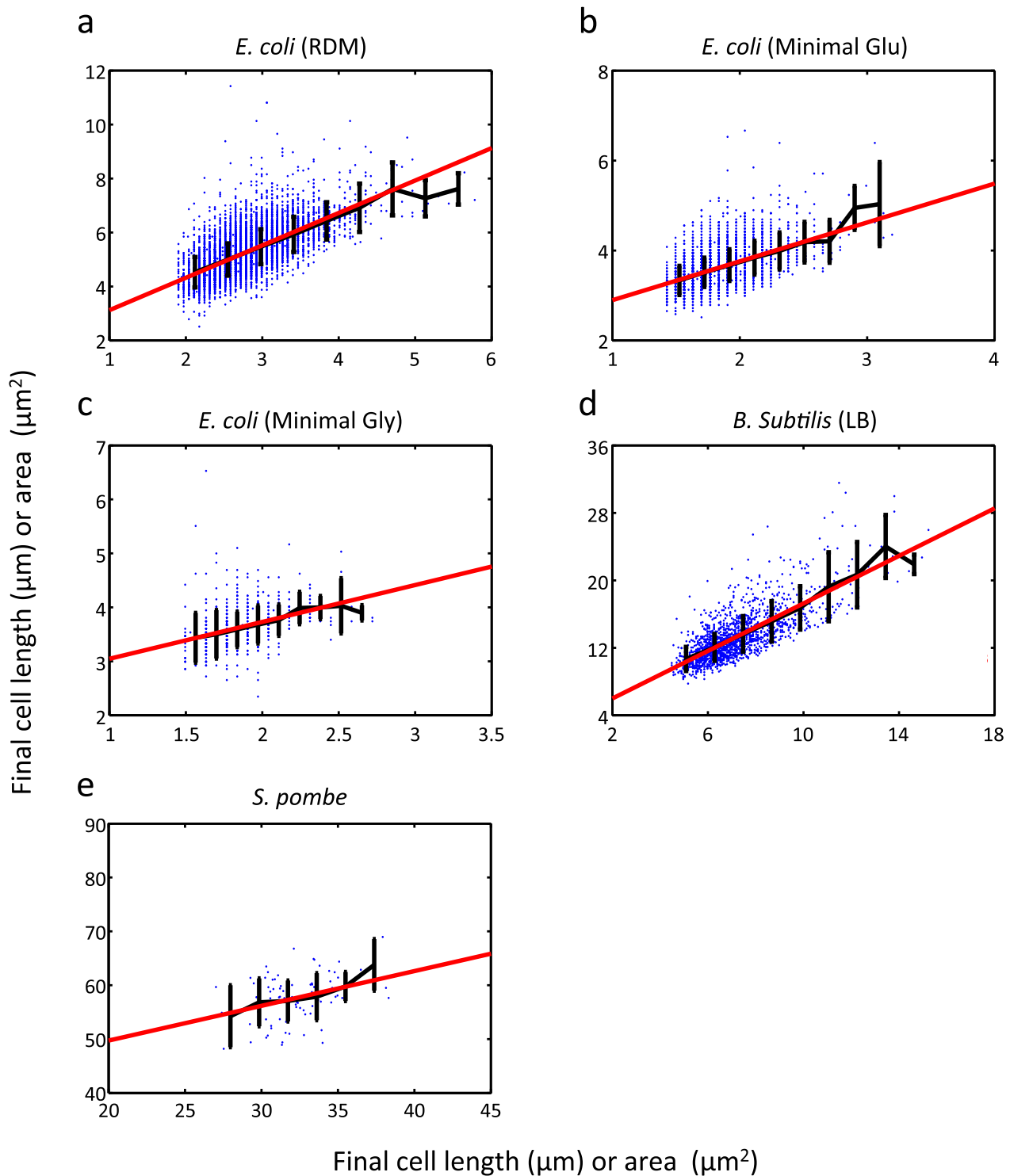
Extended Data Figure 7 | Oscillation in [Y] observed in MC4100 at 25 °C. The same plots as in Extended Data Fig. 6 but for MC4100 grown at 25 °C. $n = 61$ in c.



Extended Data Figure 8 | Oscillation in [Y] observed in MG1655. The same plots as in Extended Data Fig. 6 but for MG1655 (ref. 3). $n = 60$ in c.



Extended Data Figure 9 | Oscillation in [Y] observed in B/r strain. The same plot as in Extended Data Fig. 6 but for B/r strain³. $n = 20$ in c.



Extended Data Figure 10 | A noisy linear map in cell size control in the data sets from ref. 12 and our experimental data of *S. pombe*. The same analysis as in Fig. 1e (L_I versus L_F) was performed. These data sets were obtained using a microfluidic device different from the mother machine. **a–c**, *E. coli* growth under three different media as indicated ($n = 8,795$, $4,637$ and 684 ,

respectively). **d, e**, Growth of *B. subtilis* ($n = 1,592$) (**d**) and of *S. pombe* ($n = 85$) (**e**). Black and red lines show binned average and the linear regression line, respectively. In **a–e**, $a = 1.20$, 0.864 , 0.684 , 1.41 and 0.645 , respectively. Cell length (μm) (**a–d**) and cell area (μm^2) (**e**) are shown. The error bars indicate standard deviation of each bin.

Intersecting transcription networks constrain gene regulatory evolution

Trevor R. Sorrells^{1,2}, Lauren N. Booth^{1,2†}, Brian B. Tuch^{1,3†} & Alexander D. Johnson^{1,2}

Epistasis—the non-additive interactions between different genetic loci—constrains evolutionary pathways, blocking some and permitting others^{1–8}. For biological networks such as transcription circuits, the nature of these constraints and their consequences are largely unknown. Here we describe the evolutionary pathways of a transcription network that controls the response to mating pheromone in yeast⁹. A component of this network, the transcription regulator Ste12, has evolved two different modes of binding to a set of its target genes. In one group of species, Ste12 binds to specific DNA binding sites, while in another lineage it occupies DNA indirectly, relying on a second transcription regulator to recognize DNA. We show, through the construction of various possible evolutionary intermediates, that evolution of the direct mode of DNA binding was not directly accessible to the ancestor. Instead, it was contingent on a lineage-specific change to an overlapping transcription network with a different

function, the specification of cell type. These results show that analysing and predicting the evolution of *cis*-regulatory regions requires an understanding of their positions in overlapping networks, as this placement constrains the available evolutionary pathways.

Pathways of evolution are highly dependent on their starting points, a phenomenon often referred to as historical contingency^{1–3}. For example, studies of protein evolution have shown that initial permissive mutations are often required before a second mutation causes an evolutionary shift in function⁴. Without the initial mutation, the second mutation could be deleterious. The structural requirements for protein folding and function underlie such behaviour and are a common source of epistasis^{5–7}.

Considerably less is known about historical contingency in the evolution of transcription networks, which often comprise several transcription regulators and many target genes. It has been suggested

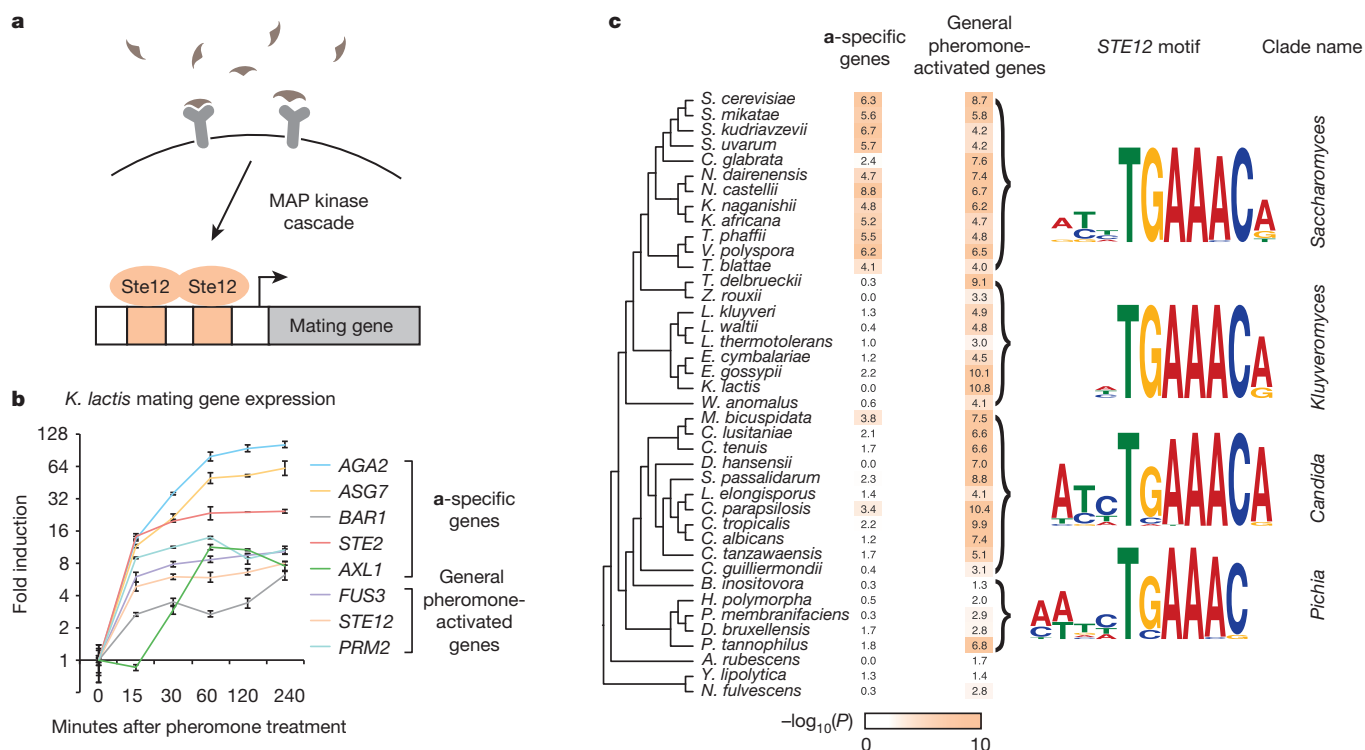


Figure 1 | Evolution of the pheromone response in budding yeast. **a**, During mating, cells sense a peptide pheromone (brown molecule) and transmit this signal to the transcription factor Ste12. **b**, *K. lactis* a cells were induced with 10 μg ml^{−1} synthetic α-factor pheromone, and gene expression was measured by quantitative reverse transcriptase PCR (qRT-PCR) over time. Shown are mean values ± s.d. for three independent genetic isolates, each grown and analysed separately. **c**, Conserved pheromone responsive genes were scored

for the presence of Ste12 *cis*-regulatory motifs in their upstream regulatory regions. The track indicates the −log₁₀(P) for the enrichment of the Ste12 motif in each set of genes in each species, by hypergeometric distribution. The genes were divided into a-specific genes and general pheromone-activated genes. The general pheromone-activated genes were used to independently generate the Ste12 *cis*-regulatory motif in each of four clades using MEME.

¹Department of Biochemistry & Biophysics, Department of Microbiology & Immunology, University of California, San Francisco, California 94158, USA. ²Tetrad Graduate Program, University of California, San Francisco, California 94158, USA. ³Biological and Medical Informatics Graduate Program, University of California, San Francisco, California 94158, USA. †Present addresses: Department of Genetics, Stanford University, Stanford, California 94305, USA (L.N.B.); Loxo Oncology, South San Francisco, California 94080, USA (B.B.T.).

that epistasis is prevalent in the evolution of regulatory networks^{8,10}, but detailed examples documenting its underlying causes are largely lacking. Here we studied the effects of contingency in the evolution of a transcription network in yeast that has diversified over several hundred million years. We found that network-level changes altered the evolutionary pathways available to individual *cis*-regulatory regions as these changes occurred at the intersection of two different but overlapping networks.

The transcription regulator Ste12 controls the response to mating pheromone in the budding yeast *Saccharomyces cerevisiae*¹¹ (Fig. 1a). When cells sense pheromone of the opposite mating type, Ste12 is activated by phosphorylation and transcriptionally upregulates many genes involved in mating¹². This function of Ste12 is conserved in the dairy yeast, *Kluyveromyces lactis*¹³ (Fig. 1b, c and Extended Data Fig. 1), a species that diverged from *S. cerevisiae* ~100 million years ago¹⁴, and in the more distantly related pathogen of humans, *Candida albicans*^{15,16}. Thus, many aspects of pheromone-activated gene expression (including the DNA sequence recognized by Ste12) have been conserved across the ~200 million years separating these three species from their common ancestor.

The genes induced by pheromone-activated Ste12 can be classified into two distinct sets, those genes specific to either the **a** or the α cell types (5–12 genes, depending on the species), and the general pheromone-activated genes (~100 genes), which are induced in both cell types. The three yeast species described above all have two mating cell types, **a** and α , that express complementary sets of genes allowing them to mate with the opposite mating type¹⁷. This study focuses on the **a**-specific genes (**asgs**), so named because they are expressed and respond to pheromone in **a** cells but not in α cells.

Using bioinformatics approaches, we quantified the number of Ste12 *cis*-regulatory sequences upstream of the **asgs** and the general pheromone-activated genes in the genomes of 40 yeasts. For the general pheromone-activated genes, Ste12 binding sites were significantly enriched (relative to the rest of the genome) across the *Candida*, *Kluyveromyces*, and *Saccharomyces* clades. Given the central, conserved role of Ste12 in the pheromone response, this enrichment was expected. However, the **asgs** showed enrichment for Ste12 binding sites only in the *Saccharomyces* clade (Fig. 1c and Extended Data Fig. 2). Although the **asgs** respond to pheromone across all three clades, they seemed to lack Ste12 *cis*-regulatory sequences in the *Kluyveromyces* and *Candida* clades. We considered three scenarios for how Ste12 activates the **asgs** in clades where these genes lack Ste12 *cis*-regulatory sequences: (1) Ste12 could bind to DNA sites in the **asgs** that are sufficiently degenerate that they fall below our threshold of detection; (2) Ste12 could be recruited through protein–protein interactions with a second transcription regulator¹⁸ that binds directly to other *cis*-regulatory sites upstream of the **asgs**; or (3) Ste12 could act ‘upstream’ by activating another regulator that binds to and activates the **asgs** directly.

We distinguished among these models (or a combination of them) by identifying the regulatory regions bound by Ste12 in *K. lactis* by chromatin immunoprecipitation followed by sequencing (ChIP-seq). As in *S. cerevisiae*⁹, *K. lactis* Ste12 was bound to the upstream regions of the **asgs** (Fig. 2a, b) as well as to the general pheromone-activated genes (Fig. 2c). Thus, *K. lactis* Ste12 is recruited to the promoters of the **asgs** and activates them in response to pheromone in the absence of recognizable Ste12 binding sites.

We considered the possibility that Ste12 is recruited to the **asgs** in *K. lactis* through a series of low-affinity *cis*-regulatory sequences or “mismatched” sequences, a situation that has been documented for certain other transcription regulators^{19,20}. In our case, the mismatched Ste12 sites are ubiquitous across the *K. lactis* genome (Extended Data Fig. 3); by mutating these mismatched sites, we showed they are not required for induction of the **asgs** by Ste12 (Fig. 2d, e; Extended Data Fig. 4).

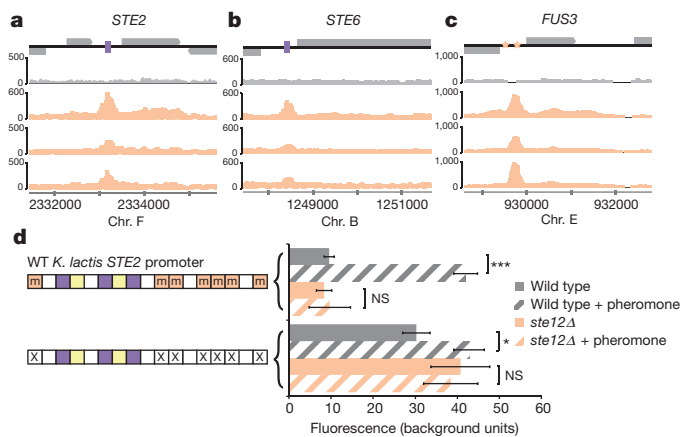


Figure 2 | Ste12 is recruited to *asg* promoters in *K. lactis* without its binding site. **a–c**, Chromatin immunoprecipitation of Ste12–myc using anti-myc antibodies (orange tracks) and an untagged control (grey track). The y axis indicates sequencing read coverage. Matches to the Ste12 binding site are shown as orange stars, and matches to the a2-Mcm1 binding site are shown as purple rectangles. Shown are two conserved **a**-specific genes (**a**, **b**), one general pheromone-activated gene (**c**). **d**, At left, organization of the *K. lactis* STE2 regulatory region. *cis*-regulatory sites for a2, Mcm1, and Ste12 are shown in purple, yellow, and orange, respectively. Sites marked with an ‘m’ reflect DNA sequences that contain a one-base-pair (bp) mismatch to the 7-bp consensus Ste12 *cis*-regulatory site. At right, the *K. lactis* STE2 regulatory region was fused to GFP and fluorescence was measured by flow cytometry. Shown is the mean fluorescence \pm s.d. for three independent genetic isolates, each grown and analysed separately. NS, not significant. Bottom, the construct contains two point mutations in each of the mismatched Ste12 *cis*-regulatory sites that should disrupt any residual Ste12 binding to these sites. Although this construct is pheromone-inducible, we note that the basal expression of the *K. lactis* promoter increased when the mismatched Ste12 sites were mutated, probably through the creation of a binding site for another activator. * $P < 0.05$, *** $P < 0.001$, by analysis of variance (ANOVA) followed by Tukey’s honest significant difference; for clarity, only the most relevant relationships are shown.

Three experimental results show that, in the *K. lactis* lineage, Ste12 is indirectly recruited to the **asgs** by association with the transcription regulator **a2**. **a2** is encoded by the *MATA2* gene (also known as *MATA2*) in the **a** mating-type locus; it is thus expressed in **a** cells (but not in α cells), where it activates transcription of the **asgs**²¹. (1) The Ste12 ChIP peaks at the **asgs** were positioned over the *cis*-regulatory motif for **a2** (Fig. 2a, b). (2) Deletion of the *MATA2* gene impairs pheromone-induction of the **asgs** (Fig. 3b, c). (3) The *cis*-regulatory sequence for **a2** is sufficient to mediate pheromone induction when moved into a naive promoter. **a2** always works with a ubiquitously expressed regulator, Mcm1 (ref. 21; Extended Data Fig. 5), so we kept the *cis*-regulatory sites of these regulators together for the purpose of this experiment. The **a2**-Mcm1 site, when inserted into a test reporter, activated green fluorescent protein (GFP) in response to pheromone (Fig. 3d and Extended Data Fig. 6). No Ste12 *cis*-regulatory sites (even mismatched ones) were present in the DNA added to the reporter construct, and the reporter itself was not pheromone-inducible. ChIP-seq was performed in a *K. lactis* strain that contained this reporter, and the **a2**-Mcm1 site was sufficient for efficient recruitment of Ste12 in the presence of pheromone (Fig. 3e). We can rule out the possibility that Mcm1 alone recruits Ste12 from the fact that Mcm1 binds to hundreds of genes in *K. lactis* and none of these are pheromone-inducible, except those also bound by **a2** (refs 22, 23). Thus, the majority of the specificity for recruitment of Ste12 must come from **a2**. We also note that Ste12 did not bind to the promoter region of the *MATA2* gene (not shown), ruling out the possibility that during the pheromone response, Ste12 acts as an upstream activator of **a2**, which then activates the **asgs**.

We now turn to the question of how the two different recruitment modes of Ste12 (direct and indirect) arose. Prior experiments in

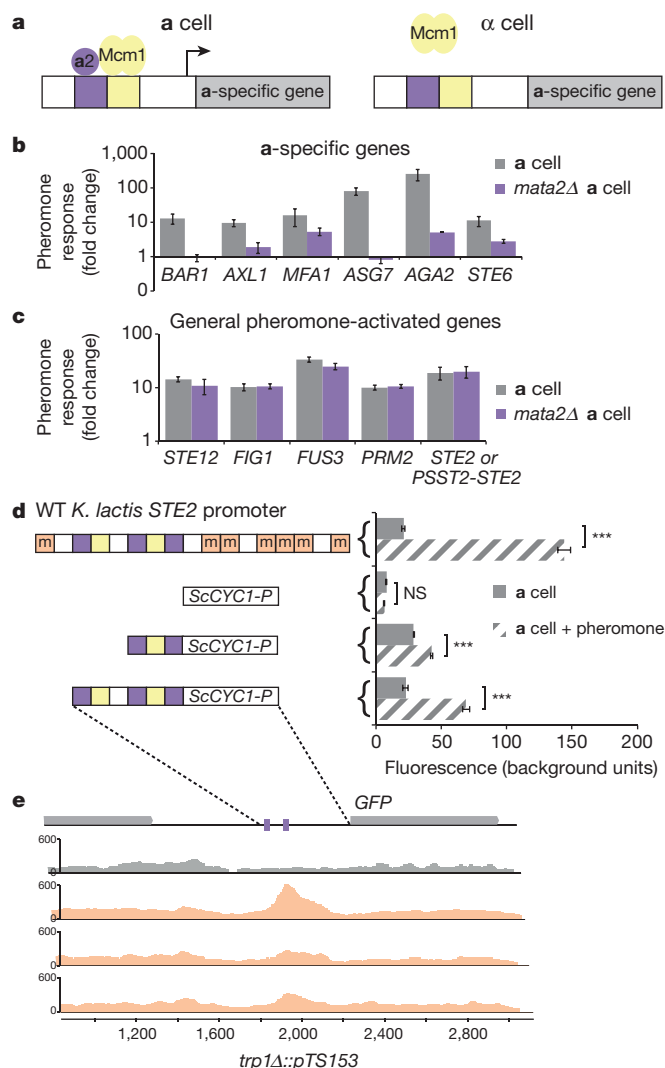


Figure 3 | a2 mediates the a-specific gene pheromone response in *K. lactis*. **a**, Diagram of asg cell-type-specific regulation in *K. lactis*. **b**, **c**, Induction of the asgs in response to pheromone in WT and *MATa2Δ* strains. Expression is shown as mean mRNA level \pm s.d. for three independent genetic isolates, each grown and analysed separately. **b**, for a-specific genes (**b**) and general pheromone-activated genes (**c**). *STE2* is shown as a general pheromone-activated gene because in this strain, its promoter was replaced by the *SST2* promoter. The differences between WT and *MATa2Δ* pheromone induction are significant in **b**, $P < 0.01$, but not in **c**, $P > 0.05$, by ANOVA. **d**, The a2-Mcm1 sites were moved into the *S. cerevisiae* *CYC1ΔUAS* promoter and expression was measured in *K. lactis* in uninduced (solid bars) and pheromone-induced (striped bars) conditions. Shown is mean fluorescence \pm s.d. of three independent genetic isolates, each grown and analysed separately. $***P < 0.001$, by ANOVA followed by Tukey's honest significant difference. **e**, Ste12-myc ChIP-Seq signal for the genome-integrated reporter construct. Symbols are described in Fig. 2.

several species^{16,21} combined with the phylogenetic distribution of Ste12 (Fig. 1c) and a2 (ref. 16) binding sites indicate that indirect recruitment by a2 at the asgs was ancestral, and that the Ste12 sites were gained in the *Saccharomyces* clade. At the general pheromone-activated genes, Ste12 is directly bound to DNA by *cis*-regulatory sequences in all three clades, providing an explanation for why Ste12 has retained the same DNA-binding specificity.

To understand why Ste12 *cis*-regulatory sites were gained at the asgs in the *Saccharomyces* clade and not other clades, we introduced these sites into an ancestral-like promoter (*STE2* of *K. lactis*) and observed the effect on asg expression (Fig. 4a). Introducing strong Ste12 sites

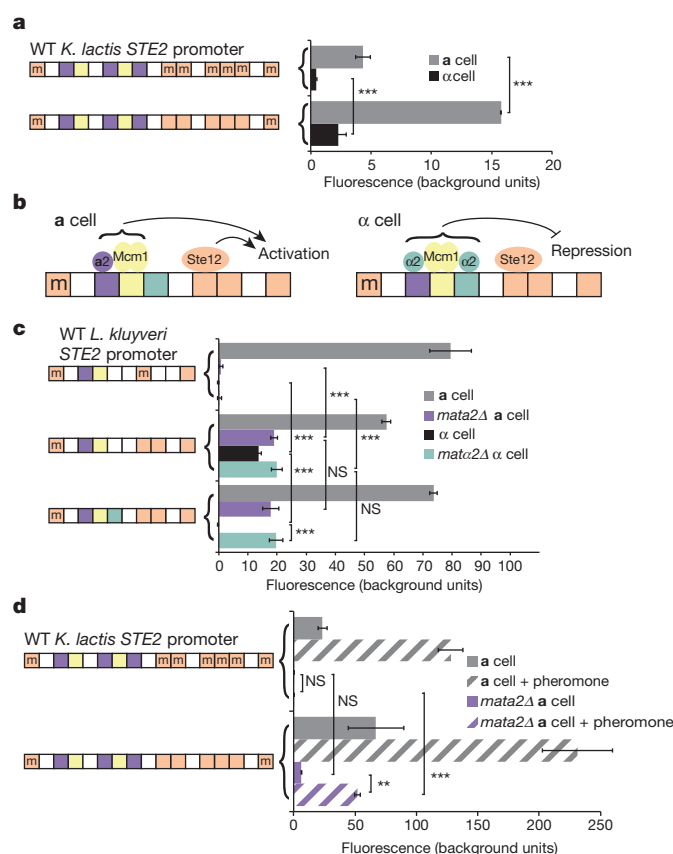


Figure 4 | Repression by a2 is necessary for the gain of Ste12 sites.

a, Expression of the *K. lactis* *STE2* reporter in the absence of pheromone. Expression is shown as mean \pm s.d. for three independent isolates, each grown and analysed separately. $***P < 0.001$; $**P < 0.01$, by ANOVA followed by Tukey's honest significant difference; for clarity, only the most relevant relationships are shown. **b**, Diagram of the *cis*-regulatory sites in the *L. kluyveri* *STE2* promoter. **c**, The *STE2* gene promoter from *L. kluyveri* was fused to GFP and integrated into the *L. kluyveri* genome. Ste12 and a2 binding sites were introduced by point mutation, and each construct was measured in four independent genetic isolates by flow cytometry. Shown is the mean fluorescence for four \pm s.d. independent isolates, each grown and analysed separately. Statistical tests shown as in **a**. **d**, Constructs with added Ste12 binding sites were tested for the ability to compensate for the deletion of the *MATa2* gene in *K. lactis*. Shown as in **a**.

by point-mutating mismatched sites in the regulatory region increased the expression in a cells, but it also resulted in mis-expression of the gene in α cells. Expression of asgs in α cells interferes with sexual reproduction because the cells respond to their own pheromones and eventually become insensitive to both a and α pheromone²⁴. Thus, we conclude—in the absence of other changes—that the evolution of high-affinity Ste12 sites disrupts the function of the ancestral form of regulation, causing misexpression of the asgs in α cells that would result in a loss of mating.

These results indicate that a prior permissive change in asg regulation was probably necessary for Ste12 sites to be gained in the *Saccharomyces* clade. In the *Saccharomyces* lineage, after the split from *Kluyveromyces*, the activator of the asgs (a2) was replaced by an α cell-type-specific repressor, $\alpha 2$ (ref. 23). To test whether this change permitted the Ste12 *cis*-regulatory sites to be gained upstream of the asgs, we performed an experiment in *Lachancea kluyveri* (Fig. 4b), a species in which all of the asgs are activated by a2, and (in contrast to *K. lactis*) two of them are also repressed by $\alpha 2$ (ref. 23). In other words, we chose a species in which both a2 and $\alpha 2$ are active. We used the *L. kluyveri* *STE2* regulatory region, which is normally regulated by a2 but not $\alpha 2$, to test the effects of introducing the Ste12

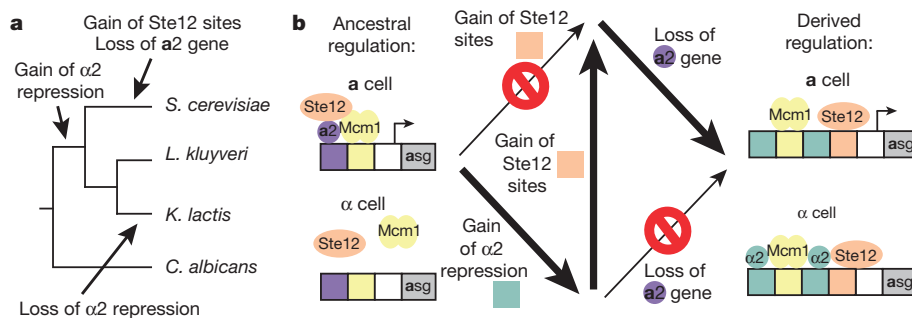


Figure 5 | Summary of key events in the evolution of *asg* regulation. **a**, The gain of $\alpha 2$ repression occurred before the gain of Ste12 *cis*-regulatory sites, whereas the loss of *MATa2* occurred on the same branch as the gain of Ste12 sites. **b**, Summary of the events that allowed and prohibited the gain of Ste12 *cis*-regulatory sequences upstream of the *a*-specific genes. Thick arrows indicate pathways that maintained regulation, and thin arrows those that would have disrupted it.

and $\alpha 2$ *cis*-regulatory sites. Introducing Ste12 *cis*-regulatory sites by point mutation again resulted in ectopic expression of the *asgs* in α cells, as we observed for *K. lactis* (Fig. 4c). However, this mis-expression could be mitigated by adding *cis*-regulatory sequences for the repressor $\alpha 2$ (Fig. 4c), the situation found naturally in *S. cerevisiae*^{24,25}. These results indicate that the switch from *a2*-mediated activation to $\alpha 2$ -mediated repression of the *asgs* was permissive for the gain of Ste12 *cis*-regulatory sequences in this lineage.

$\alpha 2$ repression was gained in the common ancestor of *Kluyveromyces* and *Saccharomyces* clades²³, and the Ste12 *cis*-regulatory sites were gained millions of years later, in the *Saccharomyces* clade (Fig. 5a). These observations indicate that, while the gain of $\alpha 2$ repression was permissive for the subsequent gain of Ste12 sites, it did not directly cause it. The gain of Ste12 sites did occur, however, at roughly the same time as another evolutionary event, the loss of the activator *MATa2* (ref. 21), suggesting that the gain of Ste12 sites may have permitted the loss of this gene (Fig. 5a). To test this possibility, we observed the expression of the *K. lactis* STE2 reporter in the absence of *a2* (Fig. 4d and Extended Data Fig. 7). Adding the Ste12 *cis*-regulatory sites could indeed compensate for the deletion of *MATa2* in both basal and pheromone-activated gene expression (Fig. 4c, d). If we assume that loss of *MATa2*—on its own—would have been deleterious, it is likely that at least some Ste12 sites were gained first, thus mitigating the potentially deleterious effects of losing *MATa2*.

Gene regulatory regions are shaped by evolutionary forces including selection and drift, each of which leave distinct patterns in the *cis*-regulatory regions of genes²⁶. To further understand the forces that shaped the evolution of *asg cis*-regulatory regions, we employed two independent approaches^{27,28}. Both methods suggested that Ste12 *cis*-regulatory sites in the *asgs* are maintained by purifying selection in the *Saccharomyces* clade, but, as predicted, the ‘mismatched’ Ste12 sites in both the *Saccharomyces* and *Kluyveromyces* clades are not (Extended Data Figs 8, 9). In particular, in the *Saccharomyces* clade, Ste12 sites were conserved with respect to the rest of the intergenic regions ($P = 1.40 \times 10^{-6}$ by likelihood ratio test (LRT)), while mismatched sites were not ($P = 1$ by LRT). In the *Kluyveromyces* clade, neither Ste12 sites ($P = 0.26$ by LRT) nor mismatched sites ($P = 1$ by LRT) show significant conservation in the *asgs*.

Our previous studies have shown how the transcription network controlling cell-type identity in yeast evolved^{16,21,23}. Here, we have shown how the evolutionary history of this network affected the evolution of an overlapping but distinct transcription network controlling the response to mating pheromone. The two networks overlap by virtue of sharing a common set of target genes. The genes at the intersection of these two networks (the *asgs*) were constrained in the evolutionary pathways available to them. In the ancestral state, the gain of Ste12 *cis*-regulatory sites were not allowed, as this would lead to cell-type misexpression. The evolution of $\alpha 2$ repression, which occurred in the ancestor of the *S. cerevisiae* clade, made the gain of these sites possible, and the loss of the *MATa2* gene at a much later time necessitated the maintenance of the Ste12 sites and prevented a return to the ancestral mode of regulation (Fig. 5b).

In protein evolution, neutral mutations can alter the effect of subsequent mutations, suggesting that evolution depends on chance events². Although we do not know whether the gain of $\alpha 2$ repression was strictly neutral, it did not change the overall *asg* expression pattern. However, it did permit changes in the overlapping circuit responsible for pheromone induction. Stated another way, changes in the *asg* circuitry altered the effect of introducing Ste12 sites to the *asgs*. Such epistatic interactions are the source of the historical contingency we observed in the *asg* regulatory network, effects that are well documented for individual proteins⁴.

Studies of enhancers across species have indicated that many molecular solutions are possible for a given output^{16,23,29,30}. Our results show that not all solutions are accessible, and those that are accessible are contingent on the functional interactions between overlapping transcription networks. These constraints are not apparent from observation of the individual components of networks (for example, a single enhancer), but only arise from complex interactions between networks. We propose that understanding the evolution of *cis*-regulatory regions (and predicting their behaviour) requires an understanding of the network structures in which they function.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 February; accepted 4 June 2015.

Published online 8 July 2015.

- Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 7899–7906 (2008).
- Harms, M. J. & Thornton, J. W. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* **512**, 203–207 (2014).
- Mohrig, J. R. *et al.* Importance of historical contingency in the stereochemistry of hydratase-dehydratase enzymes. *Science* **269**, 527–529 (1995).
- Harms, M. J. & Thornton, J. W. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nature Rev. Genet.* **14**, 559–571 (2013).
- Ortlund, E. A., Bridgman, J. T., Redinbo, M. R. & Thornton, J. W. Crystal structure of an ancient protein: evolution by conformational epistasis. *Science* **317**, 1544–1548 (2007).
- Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**, 929–932 (2006).
- Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc. Natl Acad. Sci. USA* **103**, 5869–5874 (2006).
- Gerke, J., Lorenz, K. & Cohen, B. Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323**, 498–501 (2009).
- Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
- Payne, J. L. & Wagner, A. Constraint and contingency in multifunctional gene regulatory circuits. *PLOS Comput. Biol.* **9**, e1003071 (2013).
- Dolan, J. W., Kirkman, C. & Fields, S. The yeast STE12 protein binds to the DNA sequence mediating pheromone induction. *Proc. Natl Acad. Sci. USA* **86**, 5703–5707 (1989).
- Herskowitz, I. MAP kinase pathways in yeast: for mating and more. *Cell* **80**, 187–197 (1995).
- Coria, R. *et al.* The pheromone response pathway of *Kluyveromyces lactis*. *FEMS Yeast Res.* **6**, 336–344 (2006).
- Kenschke, P. R., Oti, M., Dutilh, B. E. & Huynen, M. A. Conservation of divergent transcription in fungi. *Trends Genet.* **24**, 207–211 (2008).
- Bennett, R. J., Uhl, M. A., Miller, M. G. & Johnson, A. D. Identification and characterization of a *Candida albicans* mating pheromone. *Mol. Cell. Biol.* **23**, 8189–8201 (2003).

16. Tsong, A. E., Tuch, B. B., Li, H. & Johnson, A. D. Evolution of alternative transcriptional circuits with identical logic. *Nature* **443**, 415–420 (2006).
17. Herskowitz, I. A regulatory hierarchy for cell specialization in yeast. *Nature* **342**, 749–757 (1989).
18. Borneman, A. R. *et al.* Divergence of transcription factor binding sites across related yeast species. *Science* **317**, 815–819 (2007).
19. Ramos, A. I. & Barolo, S. Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Phil. Trans. R. Soc. Lond. B* **368**, 20130018 (2013).
20. Crocker, J. *et al.* Low affinity binding site clusters confer Hox specificity and regulatory robustness. *Cell* **160**, 191–203 (2015).
21. Tsong, A. E., Miller, M. G., Raisner, R. M. & Johnson, A. D. Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* **115**, 389–399 (2003).
22. Tuch, B. B., Galgoczy, D. J., Hernday, A. D., Li, H. & Johnson, A. D. The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* **6**, e38 (2008).
23. Baker, C. R., Booth, L. N., Sorrells, T. R. & Johnson, A. D. Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* **151**, 80–95 (2012).
24. Strathern, J., Hicks, J. & Herskowitz, I. Control of cell type in yeast by the mating type locus. The $\alpha 1$ - $\alpha 2$ hypothesis. *J. Mol. Biol.* **147**, 357–372 (1981).
25. Galgoczy, D. J. *et al.* Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **101**, 18069–18074 (2004).
26. Wray, G. A. *et al.* The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**, 1377–1419 (2003).
27. Otto, W. *et al.* Measuring transcription factor-binding site turnover: a maximum likelihood approach using phylogenies. *Genome Biol. Evol.* **1**, 85–98 (2009).
28. Hubisz, M. J., Pollard, K. S. & Siepel, A. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011).
29. Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R. & Eisen, M. B. Sepsid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLoS Genet.* **4**, e1000106 (2008).
30. Swanson, C. I., Schwimmer, D. B. & Barolo, S. Rapid evolutionary rewiring of a structurally constrained eye enhancer. *Curr. Biol.* **21**, 1186–1196 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Coyle, I. Nosedal, C. Dalal, C. Britton, C. Baker, and V. Hanson-Smith for valuable comments on the manuscript. We also thank K. Pollard, E. Mancera, C. Nobile, and Q. Mitrovich for advice and assistance with experiments and analysis. We thank K. Boundy-Mills of the Phaff Yeast Culture Collection, University of California, Davis for the *K. lactis* isolates. The work was supported by grant R01 GM037049 from the National Institutes of Health. T.R.S. was supported by a Graduate Research Fellowship from the National Science Foundation.

Author Contributions T.R.S. performed gene expression experiments, reporter assays, ChIP-Seq, and bioinformatics analyses. L.N.B. obtained and sequenced the *asg* promoters of the *K. lactis* isolates. B.B.T. conceived of, designed, and performed a preliminary analysis of the phylogenetic distribution of Ste12 sites across the *asgs*. T.R.S., L.N.B., and A.D.J. designed and interpreted experiments and edited the manuscript. T.R.S. and A.D.J. wrote the manuscript.

Author Information ChIP-Seq data has been deposited at the Gene Expression Omnibus (GEO) repository under accession number GSE65792. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.D.J. (ajohnson@cgl.ucsf.edu).

METHODS

The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Distribution of Ste12 *cis*-regulatory sites. Conserved, highly pheromone-induced genes were identified by comparing previously published expression data from *S. cerevisiae* and *C. albicans*^{15,31}. We combined results from three independent orthologue maps^{22,32,33} to find pheromone response genes that were present in most hemiascomycete yeast species. General pheromone-activated genes were defined as genes that were among the top ten most highly pheromone-activated genes in *S. cerevisiae* or *C. albicans*, present in the genomes of most hemiascomycete species, and are not an *asg* in *S. cerevisiae*, *K. lactis*, or *C. albicans*. The *asgs* were defined as showing differential expression between *a* and α cells in *S. cerevisiae*, *K. lactis*, or *C. albicans*. The protein sequences of these genes were used as a psi-blast query to identify orthologues in species whose genomes were not included in any of the orthologue maps. *MFA1* genes, which encode for the *a*-factor pheromone, are too short to be annotated by automated bioinformatics techniques, so they were identified manually in most species using tblastn.

Ste12 motifs were generated from general pheromone response gene intergenic regions pooled from all the species in a given clade. Each intergenic region was considered up to 600 bp upstream of the start codon, a distance that is likely to capture functional Ste12 *cis*-regulatory sites³⁴. The program MEME was used to find overrepresented sequences in these sets of intergenic regions, under conditions assuming zero to one binding site per intergenic region. Because the motifs generated for Ste12 from the *Kluyveromyces*, *Saccharomyces*, and *Candida* clades were highly similar, we pooled intergenic regions from all three clades and generated a 7bp motif to score individual gene regulatory regions across the hemiascomycetes. The upstream 600 bp of intergenic regions were scored as described³⁵ for the number of Ste12 motifs above a given cutoff. For 'consensus' sites, the cutoff only accepted a single sequence: TGAAACA. For mismatched sites, the cutoff allowed deviation from the consensus sequence of a single nucleotide in any position in the motif. For genes that were duplicated in a species (that is, contained more than one copy) the copy with the largest number of Ste12 sites was taken, because this was the most stringent criterion to test our hypothesis that the *a*-specific genes lacked Ste12 sites in the *Kluyveromyces* and *Candida* clades. *MFA1* genes often are found in multiple copies in each species (that is, *MFA1* and *MFA2*) so two copies were included when present.

Ste12 motifs were also enumerated in the rest of the intergenic regions in each species. Enrichment of the Ste12 motif in either the *a*-specific genes or the general pheromone response genes with respect to the rest of the genome was calculated using the hypergeometric distribution at all possible cut-offs of number of Ste12 motifs. The cutoff giving the most significant *p*-value was chosen, because this represented the most stringent test of our hypothesis that the *a*-specific genes lacked Ste12 sites in the *Kluyveromyces* and *Candida* clades.

Sequence data were obtained from the Yeast Gene Order Browser (YGOB) and the Joint Genomes Institute (JGI) websites (<http://ygob.ucd.ie> and <http://genome.jgi.doe.gov>, respectively).

Code availability. Python code used for the analysis of regulator binding sites across yeast species is available on request.

Strain construction. Constructs for gene disruption, epitope tagging, and promoter replacement were generated by fusion PCR³⁶. Three to five individual PCR components were amplified using ExTaq (Takara), and 50 ng of each were combined into a single 50 μ l reaction along with 0.5 μ l ExTaq, 0.4 μ l 25 mM dNTPs, 0.1 μ l of each 100 μ M primer. The reactions were incubated in a thermocycler: 95 °C 3:00, (95 °C 0:30, 55 °C 0:30, 72 °C 1:00 per kb) \times 4, (95 °C 0:30, 58–72 °C 0:30 (depending on the annealing temperature of the primers), 72 °C 1:00 per kb) \times 30, 72 °C 10:00.

The GFP reporters for *K. lactis*, *L. kluyveri*, and *S. cerevisiae* were constructed from the original vector that was made for use in *K. lactis*. pTS12 was made from a 5-piece fusion PCR, digested with KasI and HindIII (New England Biolabs) and ligated into pUC19 using Fast-Link ligase (Epicentre Biotechnologies). Vectors were treated with Antarctic phosphatase (New England Biolabs) for 30 min before ligation. The reporter consisted of the upstream flanking region of the *K. lactis* *TRP1* gene, the *S. cerevisiae* *TRP1* gene, the entire *S. cerevisiae* *CYC1* upstream intergenic region, the caGFP and Act1 terminator³⁷, and the downstream flanking region of *K. lactis* *TRP1*. The *S. cerevisiae* *TRP1* gene was replaced with the hygromycin resistance marker from pFA6-HTB-hphMX4³⁸ by digestion with BsiWI and PmlI and ligated to make pTS16. The *CYC1* promoter UAS was deleted by digestion with XhoI. DNA oligonucleotides with sites for NotI and BamHI sites were annealed in Fast-link ligase buffer (Epicentre Biotechnologies) and ligated in a 50:1 insert to vector ratio to make pTS26. This vector was used to make the other reporters for use in *K. lactis* by ligating annealed oligonucleotides or digested PCR products into the NotI and XhoI sites. Vectors incorporating an entire intergenic

region of a gene were constructed by ligating digested PCR products the sites SacI and AgeI. The *L. kluyveri* reporter was made by cloning upstream and downstream homology to the *TRP1* locus into the NgoMIV/BstEII and BlnI/HindIII sites, respectively. These homology regions were then swapped out for homology to the *URA3* locus to improve the efficiency of integration of the reporter by selection with 5-FOA. The altered regions of the *L. kluyveri* *STE2* promoter were cloned as with the *K. lactis* reporter. The *S. cerevisiae* reporter was made by cloning the *STE2* promoter fused to GFP into a centromeric reporter, pRS412, using the KpnI and SacI sites.

To measure the pheromone response in a strain that lacks *MATa2*, we replaced the promoter of *STE2* with that of a general pheromone response gene that is not dependent on *a2*. Strain yTS43 in which the *STE2* promoter was replaced by the *SST2* promoter was made in the *MATa2A* background. This replacement construct consisted of a 4-piece fusion PCR that contained 5' homology, the *URA3* marker, the *SST2*-promoter, and 3' homology.

K. lactis and *L. kluyveri* were transformed with 1 μ g DNA according to published protocols^{39,40}, and *S. cerevisiae* was transformed using a standard lithium acetate protocol. Yeast were grown for 1 day on YEPD then replica plated onto selective media.

RT-qPCR. *K. lactis* cells were grown overnight for ~14 h, then starved in SD medium lacking phosphate as previously described²². Cells were induced after 6 h with 6.25 μ M α -factor pheromone (WSWTLR PGQPIF >95% purity, 100 mg ml⁻¹ in 100% DMSO; Genemed Synthesis) or an equivalent amount of DMSO. Cells were removed at several time points and were pelleted, washed, and frozen in liquid nitrogen. Mutant strains were harvested at 4 h, the final time point. Each experiment was performed in triplicate a single time. RNA was extracted using the RiboPure RNA Purification kit (Ambion). The RNA was reverse transcribed into cDNA using Superscript II as described previously⁴¹. Transcript levels were measured using SYBR green on a StepOnePlus RT PCR machine (Applied Biosystems). Transcripts were normalized to that of *VPS4*, a gene that was found to be highly consistent across conditions in previous expression experiments⁴². Statistics were performed using an ANOVA.

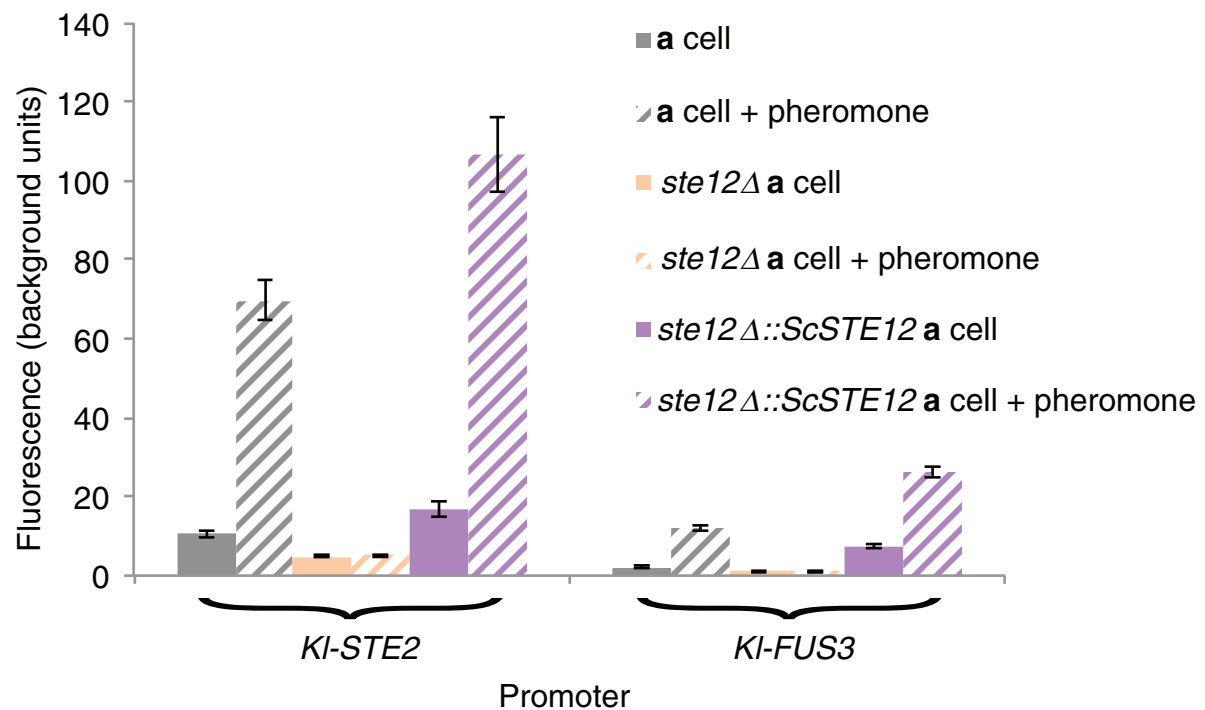
Chromatin immunoprecipitation. Strains yTS314 and yTS315 were grown using a phosphate starvation protocol modified from our previously published protocol²². Yeast were grown overnight for ~14 h, then pelleted, resuspended in phosphate starvation media, then diluted in 100–200 ml phosphate starvation media to OD₆₀₀ = 0.25–0.3. These cultures were grown for 2 h, then 13-mer α -factor pheromone was added to 6.25 μ M, and the cultures were grown for an additional 2 h. Cells were crosslinked with 1% formaldehyde for 15 min, quenched with glycine, pelleted, washed, and frozen in liquid nitrogen. Cells were lysed in 300–700 μ l lysis buffer (50 mM HEPES/KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate) with added EDTA-free protease inhibitor tablet (Roche). Cells were transferred to a fresh tube with 500 μ l 0.4 mm glass beads, and were lysed for ~45 min on a vortex genie with tube adaptor (LabRepCo). Lysate was recovered by centrifugation, and sonicated 3–4 \times on a Diagenode Bioruptor (level 5, 30 s on, 1 min off). Lysate was cleared by centrifugation at max speed in a table-top centrifuge at 4 °C. Immunoprecipitation was carried out overnight with 300 μ l cleared lysate, 200 μ l fresh lysis buffer, and 10 μ l 200 μ g ml⁻¹ anti-c-myc antibody (Invitrogen). 50 μ l of washed Protein G sepharose beads 50% slurry was added and incubated for 2 h. The beads were washed and eluted as previously described⁴³. Crosslinking was reversed by incubating 16 h at 65 °C, and immunoprecipitated DNA was recovered using the MinElute kit (Qiagen). ChIP-Seq libraries were prepared using the NEBNext ChIP-Seq Library Prep kit for Illumina (New England Biolabs) and AMPure XP magnetic beads (Beckman Coulter, Inc.). Size distribution of the libraries was determined with a BioAnalyzer 2100 (Agilent). Libraries were pooled and sequenced on a HiSeq 2500 (Illumina) through the UCSF Center for Advanced Technology (cat.ucsf.edu). Quality of reads was checked using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were indexed and aligned to the genome using Bowtie 2⁴⁴. File types were manipulated using SAMtools⁴⁵, and peaks were called using MACS⁴⁶. Coverage and peak calls were visualized in MochiView⁴⁷. The ChIP experiment was performed with a total of three replicates on two separate days, and a single replicate was performed of the untagged control.

Reporter assays. *K. lactis* reporter strains were grown as 1 ml cultures in SD medium overnight for ~14 h in a 96-well plate. Cells were pelleted, resuspended in phosphate starvation medium, and diluted to OD₆₀₀ = 0.25. Cells were grown 6 h, then induced with 6.25 μ M α -factor (or DMSO for control cells) for 4 h. Cells were diluted tenfold into the same media before measuring GFP fluorescence on a BD LSR II Flow Cytometer. A total of 10,000 cell measurements were taken for each strain. Cells were gated to exclude debris and the mean of the cell population was used for further analysis. The autofluorescence level of a cell containing no reporter was subtracted from each reporter strain, and the standard deviation of each reporter strain was added to the standard deviation of the autofluorescence

strain. Values were divided by the mean autofluorescence to give an indication of the signal relative to noise. Multiple independent isolates were tested for each reporter strain, and each experiment was performed separately two or more times. In rare cases where one isolate displayed a drastic difference in expression from the others, it was tested to see whether it was a statistical outlier, and if so, was removed for subsequent experimental repetitions. Statistics for the reporter assays were conducted using an ANOVA followed by Tukey's HSD test.

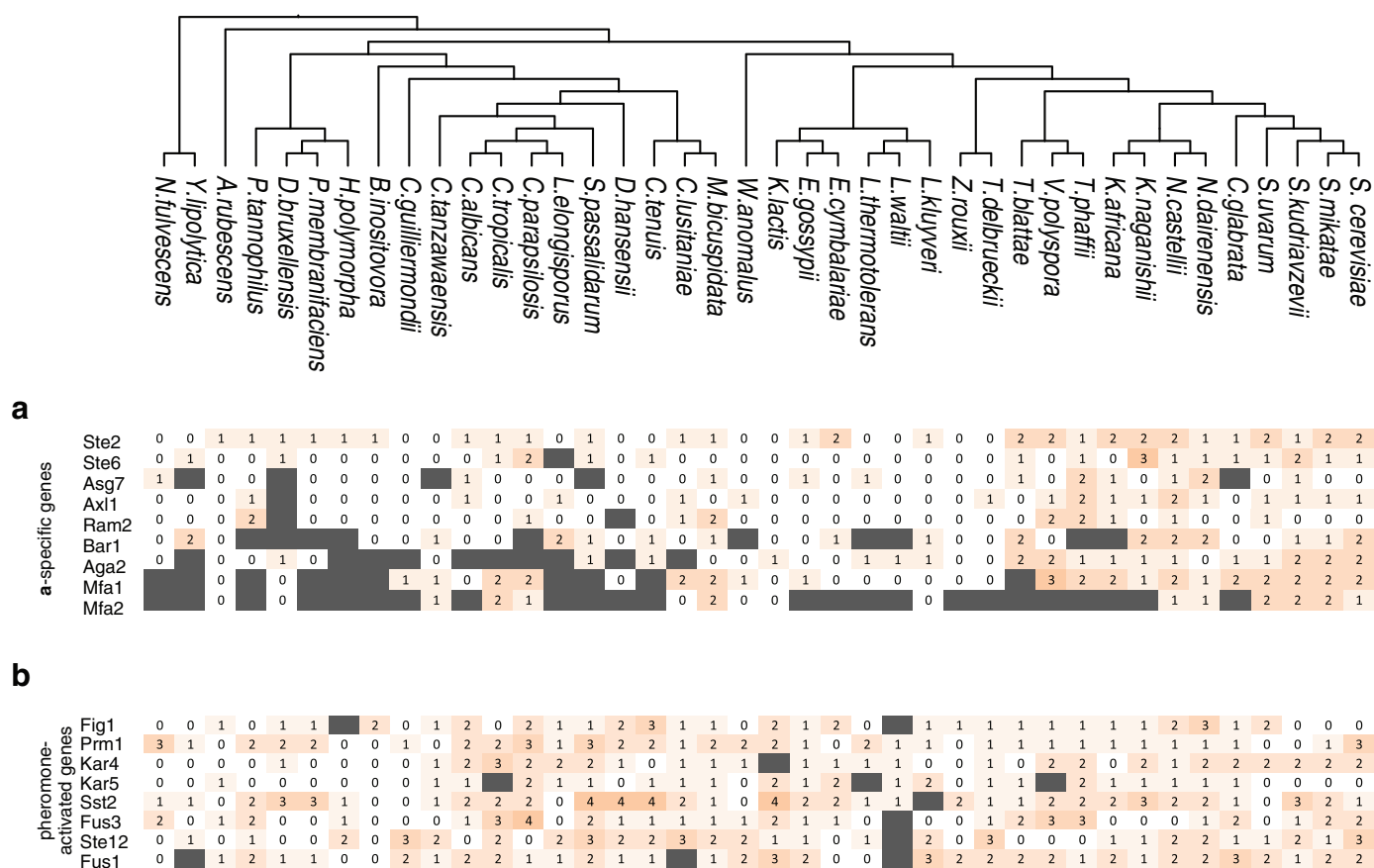
Selection on Ste12 *cis*-regulatory sites. The intergenic sequences for the *a*-specific genes were obtained from YGOB for the species *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. uvarum*, and *S. kudriavzevii*. 19 *K. lactis* strains were obtained from the Phaff Yeast Culture Collection (Davis, California). From these and 3 additional *K. lactis* strains, *a*-specific gene intergenic sequences were amplified from both directions using ExTaq (Takara) and Sanger sequenced. The intergenic regions from the *Saccharomyces sensu stricto* and *K. lactis* species complex were then analysed separately as follows. For each species or isolate, the *a*-specific gene intergenic regions were scored for the presence of strong and weak Ste12 *cis*-regulatory regions. The intergenic regions were aligned with MUSCLE⁴⁸, and concatenated for tree building. The appropriate nucleotide evolution model was selected as previously described⁴⁹; the HKY+G model was selected for the *sensu stricto* species, and the GTR+G model was selected for *K. lactis*. These models output the estimated evolutionary rate for each site in the alignment and each site was categorized as being part of a consensus or mismatched Ste12 *cis*-regulatory site in at least one species, or no site at all. Then, the trees generated by this method were used as input to test for different evolutionary rates among the different categories using the RPHAST package. The null model was generated using phyloFit on the intergenic regions lacking mismatched and consensus Ste12 sites using the 'HKY85' model for the *sensu stricto* species and the 'REV' model for *K. lactis*. Then, the mismatched and consensus Ste12 sites were tested for increased conservation in comparison the null model using phyloP.

31. Tirosh, I., Weinberger, A., Bezael, D., Kaganovich, M. & Barkai, N. On the relation between promoter divergence and gene expression evolution. *Mol. Syst. Biol.* **4**, 159 (2008).
32. Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54–61 (2007).
33. Gordon, J. L. *et al.* Evolutionary erosion of yeast sex chromosomes by mating-type switching accidents. *Proc. Natl Acad. Sci. USA* **108**, 20024–20029 (2011).
34. Lusk, R. W. & Eisen, M. B. Spatial promoter recognition signatures may enhance transcription factor specificity in yeast. *PLoS ONE* **8**, e53778 (2013).
35. D'haeseleer, P. What are DNA sequence motifs? *Nature Biotechnol.* **24**, 423–425 (2006).
36. Wach, A. PCR-synthesis of marker cassettes with long flanking homology regions for gene disruptions in *S. cerevisiae*. *Yeast* **12**, 259–265 (1996).
37. Cormack, B. P. *et al.* Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in *Candida albicans*. *Microbiology* **143**, 303–311 (1997).
38. Tagwerker, C. *et al.* HB tag modules for PCR-based gene tagging and tandem affinity purification in *Saccharomyces cerevisiae*. *Yeast* **23**, 623–632 (2006).
39. Kooistra, R., Hooykaas, P. J. J. & Steensma, H. Y. Efficient gene targeting in *Kluyveromyces lactis*. *Yeast* **21**, 781–792 (2004).
40. Gojkovic, Z., Jahnke, K., Schnackerz, K. D. & Piskur, J. PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *J. Mol. Biol.* **295**, 1073–1087 (2000).
41. Mitrovich, Q. M., Tuch, B. B., Guthrie, C. & Johnson, A. D. Computational and experimental approaches double the number of known introns in the pathogenic yeast *Candida albicans*. *Genome Res.* **17**, 492–502 (2007).
42. Booth, L. N., Tuch, B. B. & Johnson, A. D. Intercalation of a new tier of transcription regulation into an ancient circuit. *Nature* **468**, 959–963 (2010).
43. Borneman, A. R. *et al.* Target hub proteins serve as master regulators of development in yeast. *Genes Dev.* **20**, 435–448 (2006).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357–359 (2012).
45. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
46. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
47. Homann, O. R. & Johnson, A. D. MochiView: versatile software for genome browsing and DNA motif analysis. *BMC Biol.* **8**, 49 (2010).
48. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
49. Posada, D. & Crandall, K. A. Selecting models of nucleotide substitution: an application to human immunodeficiency virus 1 (HIV-1). *Mol. Biol. Evol.* **18**, 897–906 (2001).



Extended Data Figure 1 | Conservation of Ste12 function. Expression of the *asg STE2* and the general pheromone-activated gene *FUS3* in the presence of *K. lactis* Ste12, in the absence of Ste12, and in the presence of *S. cerevisiae*

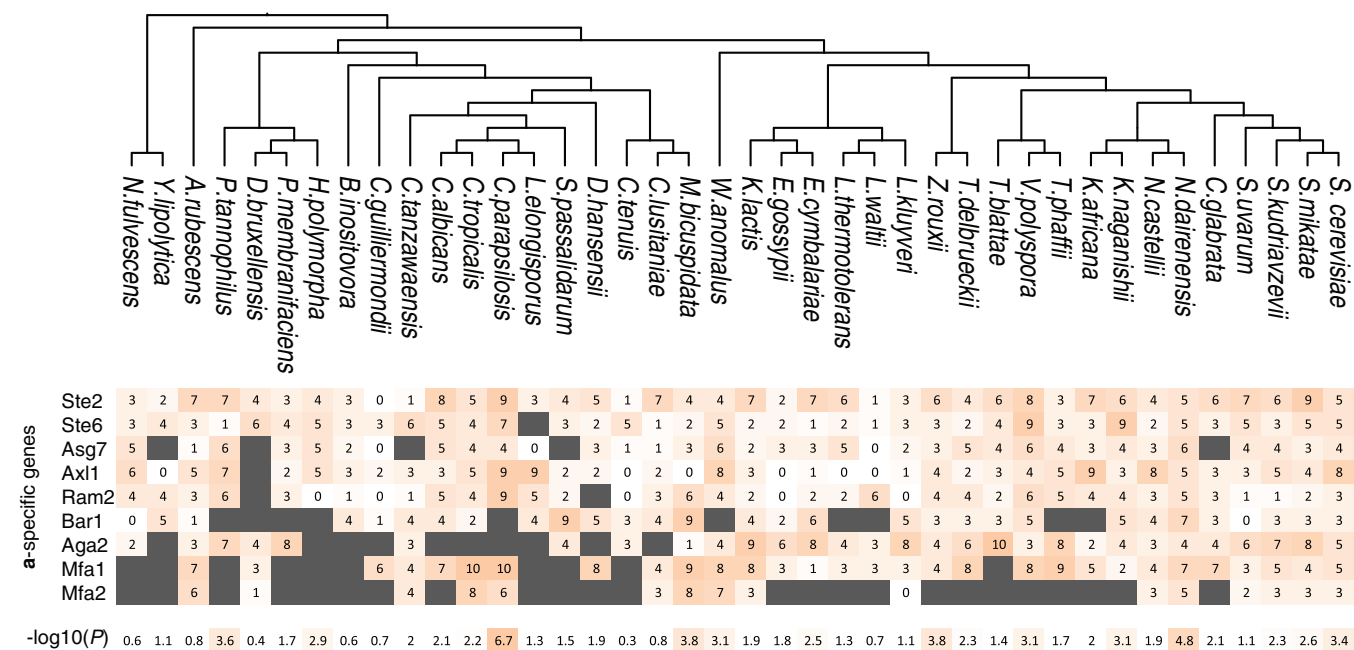
Ste12. Expression under uninduced and pheromone-induced conditions are shown as mean fluorescence \pm s.d. of three independent genetic isolates, grown and analysed separately.



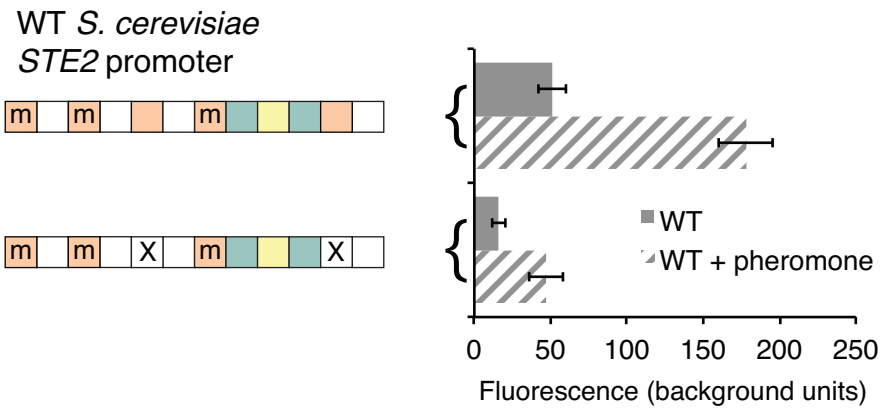
Extended Data Figure 2 | Distribution of Ste12 cis-regulatory sites.

Individual a-specific genes and general pheromone-activated genes were scored for the presence of Ste12 cis-regulatory motifs in their upstream regulatory regions. Shown is the number of consensus Ste12 sites within 600 bp of the

translation start site. Orthologues that could not be identified in a given species are shown in grey. The enrichment for the Ste12 motif in each set of genes in each species is shown in Fig. 1c.

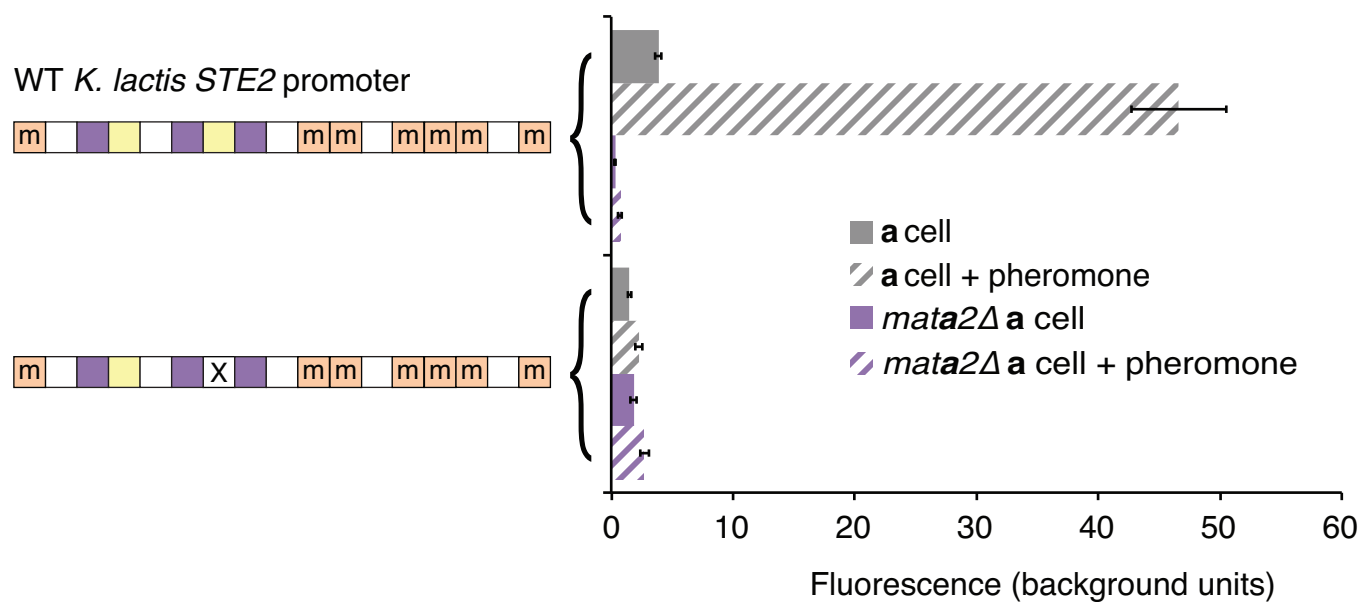


Extended Data Figure 3 | Distribution of mismatched Ste12 cis-regulatory sites. asgs were scored for the presence of mismatched Ste12 binding sites in their upstream regulatory regions. The enrichment of the Ste12 site in the asgs in each species is shown in the bottom row.

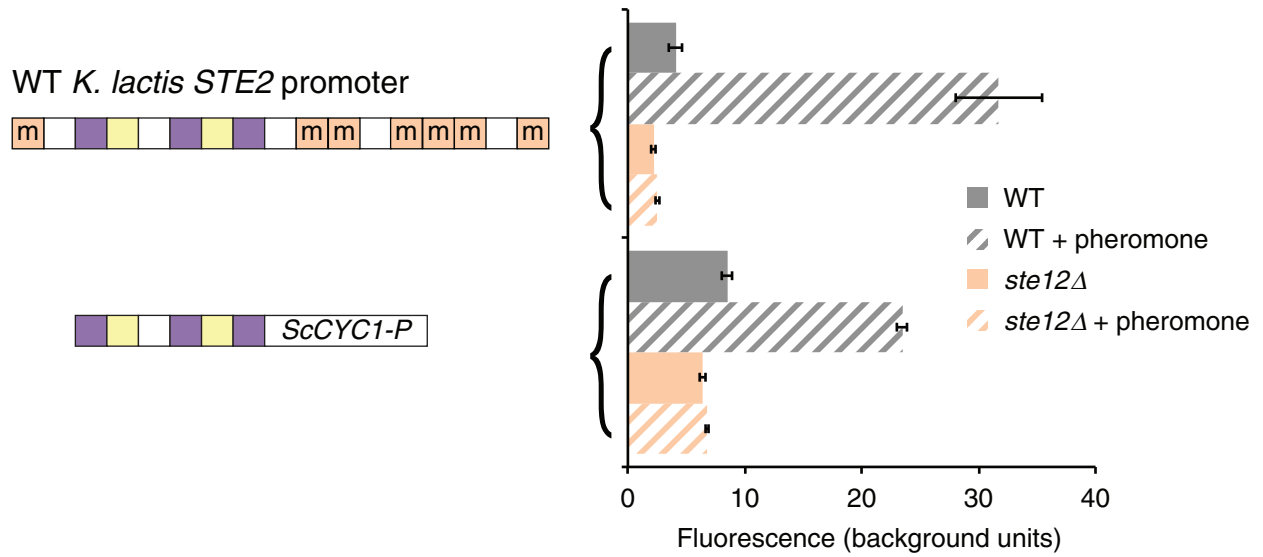


Extended Data Figure 4 | Ste12 sites are required in *S. cerevisiae* asgs. The *S. cerevisiae* STE2 promoter was fused to GFP and the role of Ste12 *cis*-regulatory sites in expression. Expression under uninduced and

pheromone-induced conditions are shown as mean fluorescence \pm s.d. of three independent genetic isolates, grown and analysed separately. Symbols are described in Fig. 2d.

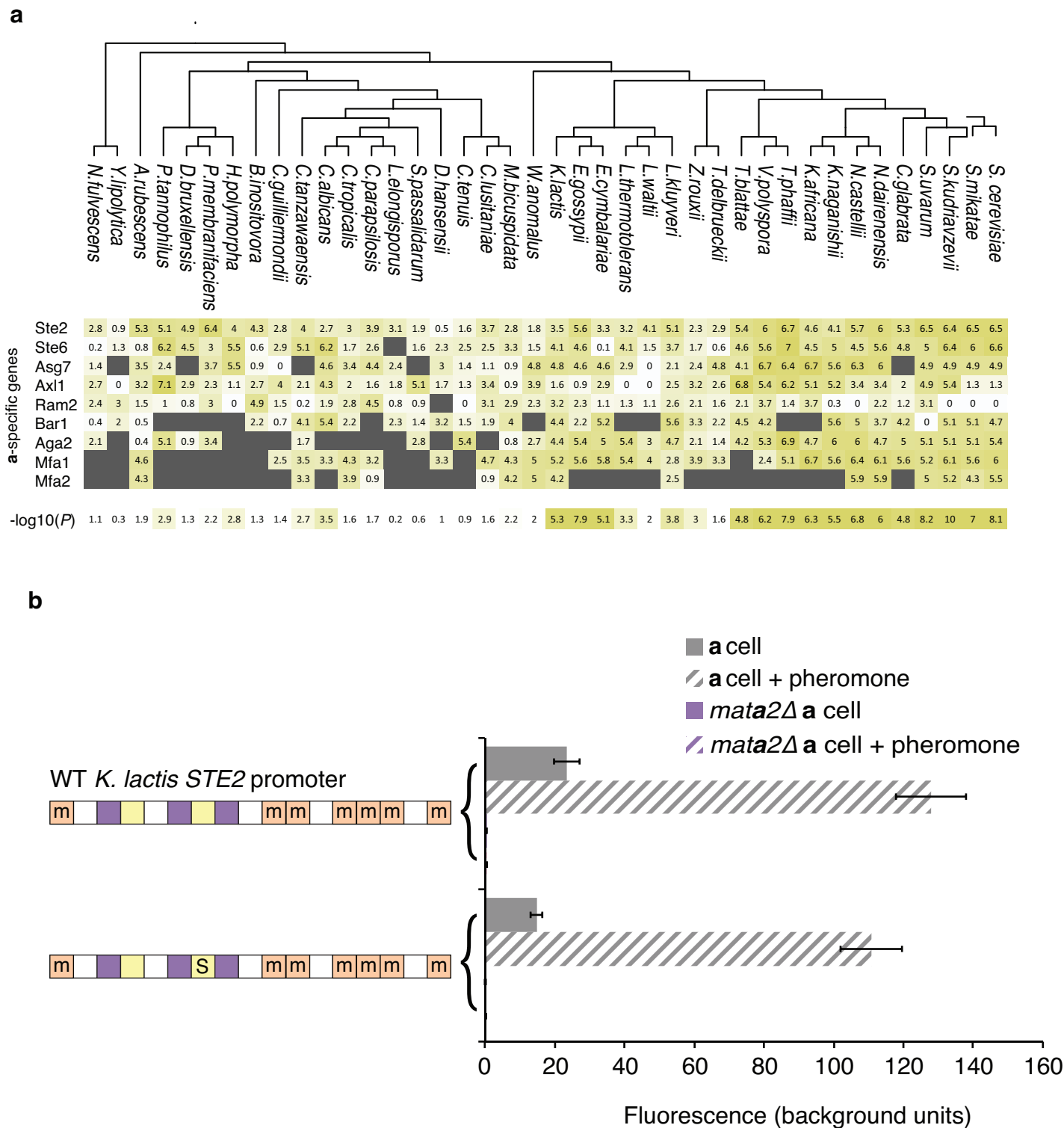


Extended Data Figure 5 | Mcm1 sites are required in *K. lactis* *asgs*. The *K. lactis* *STE2* GFP reporter was tested with a mutated Mcm1 *cis*-regulatory site. Shown is mean fluorescence \pm s.d. of three independent genetic isolates, grown and analysed separately. Symbols are described in Fig. 3d.



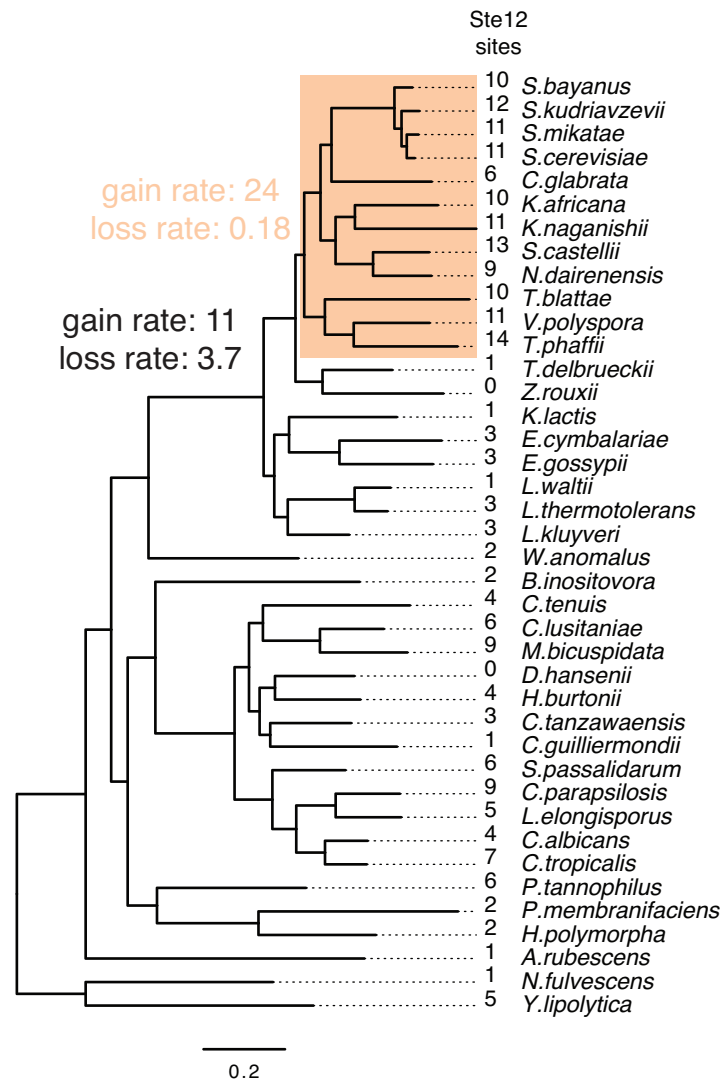
Extended Data Figure 6 | a2 sites are sufficient for pheromone activation. The *K. lactis* STE2 GFP reporter and heterologous reporter containing the a2-Mcm1 binding site were transformed into wild type and *ste12Δ* cells.

Shown is mean fluorescence \pm s.d. of three independent genetic isolates, grown and analysed separately.



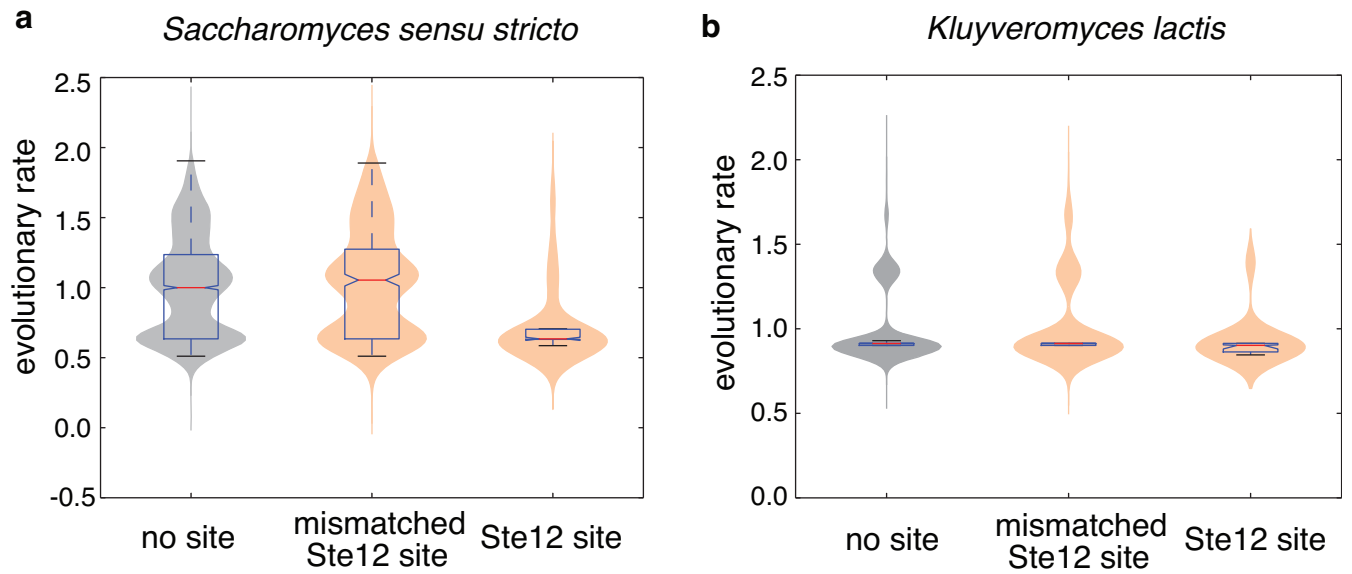
Extended Data Figure 7 | Mcm1 sites do not compensate for the loss of a2.
a. The presence of the Mcm1 binding site is unchanged across the *Candida*, *Kluyveromyces*, and *Saccharomyces* clades^{22,24}. *asg*s were scored for the strength of the Mcm1 binding site in their upstream regulatory regions. The enrichment of the strength of the Mcm1 site in the *asg*s in each species is shown in the

bottom row. **b.** A construct with increased Mcm1 binding site strength, marked with an 'S', was tested for its ability to compensate for the deletion of *a2*. Shown is the mean fluorescence \pm s.d. of three independent genetic isolates, grown and analysed separately.



Extended Data Figure 8 | Loss of Ste12 sites decreased in the *Saccharomyces* clade. The yeast phylogeny along with the number of Ste12 binding sites in extant species was used to estimate the gain and loss rate of the sites over evolutionary time. The *Saccharomyces* clade was allowed to have different

rates (orange) (LRT, $X^2 = 49.33$, $df = 2$, $P = 1.94 \times 10^{-11}$). The gain and loss rate units are *cis*-regulatory sites per amino acid substitution in the species phylogeny.



Extended Data Figure 9 | *Saccharomyces* Ste12 binding sites evolve under purifying selection. The evolutionary rate of nucleotides in Ste12 binding sites was compared to rates in the rest of the upstream regulatory regions of the

a-specific genes, shown as violin plots. Promoters between closely related species were aligned and the evolutionary rate of each base pair in the alignment was determined after model selection.

Structural basis for retroviral integration into nucleosomes

Daniel P. Maskell¹, Ludovic Renault^{2,3}, Erik Serrao⁴, Paul Lesbats¹, Rishi Matadeen⁵, Stephen Hare^{6†}, Dirk Lindemann⁷, Alan N. Engelman⁴, Alessandro Costa² & Peter Cherepanov^{1,6}

Retroviral integration is catalysed by a tetramer of integrase (IN) assembled on viral DNA ends in a stable complex, known as the intasome^{1,2}. How the intasome interfaces with chromosomal DNA, which exists in the form of nucleosomal arrays, is currently unknown. Here we show that the prototype foamy virus (PFV) intasome is proficient at stable capture of nucleosomes as targets for integration. Single-particle cryo-electron microscopy reveals a multivalent intasome–nucleosome interface involving both gyres of nucleosomal DNA and one H2A–H2B heterodimer. While the histone octamer remains intact, the DNA is lifted from the surface of the H2A–H2B heterodimer to allow integration at strongly preferred superhelix location ± 3.5 positions. Amino acid substitutions disrupting these contacts impinge on the ability of the intasome to engage nucleosomes *in vitro* and redistribute viral integration sites on the genomic scale. Our findings elucidate the molecular basis for nucleosome capture by the viral DNA recombination machinery and the underlying nucleosome plasticity that allows integration.

Retroviral INs and related transposases severely deform target DNA (tDNA) to gain access to the scissile phosphodiester bonds^{3,4}. Given the limited accessibility and constraints imposed on the conformation of DNA, the nucleosomal structure⁵ should be expected to impede integration. Yet mounting evidence indicates that retroviruses and some yeast retrotransposons integrate into nucleosomes^{6–12}. Recombinant PFV IN enables assembly of all the key intermediates of retroviral integration^{2,3,13,14}, presenting opportunities for hitherto unprecedented experimental approaches to probe interactions between the viral machinery and its cellular partners. The PFV intasome displayed robust strand transfer activity when supplied with mononucleosomes prepared from human chromatin or the previously described recombinant W601 nucleosome¹⁵. The reaction yielded two major types of DNA products (long (L) and short (S)), consistent with concerted integration into the exposed major groove at nucleosomal superhelix location (SHL) ± 3.5 positions, separated from the dyad by 3.5 turns of DNA helix (~ 36 bp; Fig. 1a and Extended Data Fig. 1a–c). By contrast, integration into deproteinized nucleosomal DNA was far less efficient and lacked pronounced hotspots (Fig. 1a and Extended Data Fig. 1c). Nucleosomes could be pulled-down by biotinylated intasome on streptavidin agarose under a range of salt concentrations in the absence of divalent metal cofactors, which are essential for IN enzymatic activity. The substitution A188D in IN suppressed the interaction, confirming involvement of the intasomal tDNA-binding groove in nucleosome capture (Extended Data Fig. 2a)³.

To identify a nucleosome suitable for structural studies in complex with the intasome, we isolated human nucleosomes captured by the intasome in the presence of 290 mM NaCl (Extended Data Fig. 2b). Three individual nucleosomal DNA fragments recovered in this experiment were assembled with recombinant human histones

(Extended Data Figs 1a and 2c). While displaying the common PFV integration hotspots at SHL ± 3.5 positions (Fig. 1a and Extended Data Fig. 1d), the selected nucleosomes D02, F02 and H04 bound the intasome under considerably more stringent conditions compared with W601 (Fig. 1b), a property that depended on nucleosome structure (Extended Data Fig. 2d). Lower thermal stability of the selected nucleosomes (Extended Data Fig. 3) suggests enhanced flexibility, which may aid in the conformational adaptation required for intasome binding (see later). The D02 nucleosome afforded isolation of a stable complex with the intasome, which, upon incubation with 5 mM Mg^{2+} , converted into the strand transfer complex with integrated viral DNA ends (Fig. 1c, d). DNA sequencing analysis of the resulting products revealed integration into a single site, offset from the middle of the D02 DNA by 36 bp, indicating that the complexes comprised the dyad-related nucleosomal site dissociated during purification (Extended Data Fig. 1d).

To determine the structure of the 400 kDa intasome–D02 nucleosome complex before strand transfer, we acquired single-particle cryo-electron microscopy (EM) data. The resulting electron density map, calculated to 7.8 Å resolution (Extended Data Fig. 4), allowed unambiguous docking of the intasome^{2,3} and the nucleosome^{5,16} crystal structures (Fig. 2a). The intasome contains a homotetramer of IN made of two types of subunits. The inner IN chains provide catalytic function, synapse the viral DNA ends and form the tDNA-binding groove. The function of the outer IN subunits, which attach to the inner subunits via the canonical catalytic core domain (CCD) dimerization interface^{2,17}, has been unclear. The path of the viral and nucleosomal DNA backbone, the histone octamer and the inner subunits of the IN tetramer are well defined in the electron density map (Figs 2a, 3a and Supplementary Video 1).

The cryo-EM structure reveals an extensive intasome–nucleosome interface, involving three IN subunits, both gyres of the nucleosomal DNA and one H2A–H2B heterodimer (Fig. 2a and Supplementary Video 1). The tDNA-binding groove of the intasome engages nucleosomal DNA above one of the H2A–H2B histone heterodimers, in agreement with the location of the preferred integration site. The path of DNA captured within the tDNA-binding groove of the intasome strongly deviates from that in the structure of free nucleosome (Fig. 2b). Here, DNA is deformed and lifted by ~ 7 Å from the surface of the H2A–H2B heterodimer to adopt a conformation strikingly matching the sharply bent naked tDNA in crystals of the PFV target capture complex³ (Fig. 2b, Extended Data Fig. 5 and Supplementary Video 1). Notably, the H2A L1-loop directly underlying the integration site is a hotspot for structural variability within the histone core¹⁸. Thus, the segment of nucleosomal DNA preferentially targeted by the intasome may be particularly malleable to conformational adaptation. The interface is extended by ancillary contacts on both sides of the tDNA-binding groove. A triad of loops belonging to the

¹Chromatin Structure and Mobile DNA, The Francis Crick Institute, Blanche Lane, South Mimms EN6 3LD, UK. ²Architecture and Dynamics of Macromolecular Machines, Clare Hall Laboratories, The Francis Crick Institute, Blanche Lane, South Mimms EN6 3LD, UK. ³National Institute for Biological Standards and Control, Microscopy and Imaging, Blanche Lane, South Mimms EN6 3QG, UK. ⁴Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, 450 Brookline Avenue, Boston, Massachusetts 02215, USA. ⁵NeCCN, Gorlaeus Laboratory, Einsteinweg 55, Leiden 2333, the Netherlands. ⁶Division of Medicine, Imperial College London, St-Mary's Campus, Norfolk Place, London W2 1PG, UK. ⁷Institute of Virology, Technische Universität Dresden, Fetscherstr. 74, Dresden 01307, Germany. [†]Present address: Division of Molecular Biosciences, Imperial College London, South Kensington Campus, London SW7 2AZ, UK.

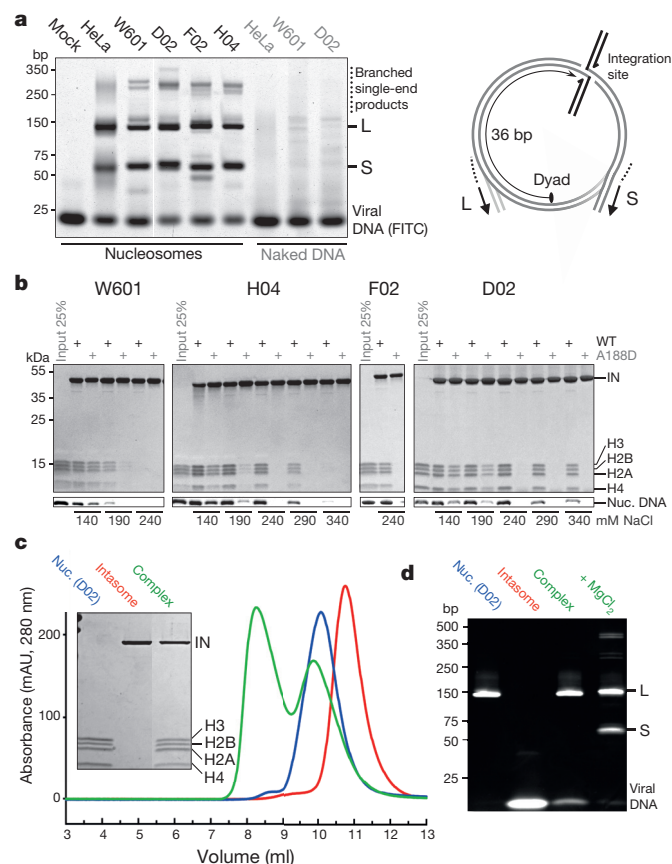


Figure 1 | Nucleosome capture by the PFV intasome. **a**, Integration into HeLa-derived or recombinant (W601, H04, D02 or F02) core nucleosomes or naked DNA. Fluorescein-labelled intasomal DNA and reaction products were separated by polyacrylamide gel electrophoresis (PAGE) and detected by fluorescence scanning. The major long (L; ~127 bp) and short (S; ~56 bp) products result from concerted strand transfer at nucleosomal SHL ± 3.5 positions, which are separated from the dyad by 3.5 turns of DNA helix or 36 bp. The products include viral DNA joined to nucleosomal DNA fragments (Extended Data Fig. 1b). FITC, fluorescein isothiocyanate. **b**, Pull-down of recombinant nucleosomes with biotinylated intasome in the presence of variable NaCl concentrations; where indicated, the intasome was assembled with A188D IN. Bound material, separated in SDS-PAGE gels, was stained to detect proteins (IN, H3, H2A, H2B and H4) and nucleosomal DNA (Nuc. DNA). WT, wild type. **c**, Isolation of the intasome–D02 nucleosome complex by size-exclusion chromatography. Peak fractions of intasome (red trace), D02 nucleosome (blue) and the complex (green) were separated by SDS-PAGE (inset). **d**, DNAs from D02 nucleosome, intasome, and the intasome–nucleosome complex, before and after incubation with 5 mM $MgCl_2$, were separated by PAGE and detected with GelRed.

inner IN subunits cradle the carboxy-terminal helix of H2B (Fig. 2a). On the other side, the intasome leans against the second gyre of nucleosomal DNA, using a saddle-shaped surface of the CCD–CCD interface (Figs 2a and 3a; secondary interface). These interactions may compensate for the requirement to deform DNA beyond its ground state on the nucleosome. An early study, albeit using undefined integrase–DNA complexes, implicated SHL ± 3.5 and SHL ± 1.5 positions as sites for human immunodeficiency virus (HIV)-1 integration into nucleosomes⁸. Thus, some of the nucleosomal locations may be virus-species dependent, perhaps directed by interactions between integrase and specific histone heterodimers.

Next, we introduced amino acid substitutions into the relevant regions of PFV IN and evaluated the ability of the resulting intasomes to engage nucleosomes *in vitro*. Some substitutions were engineered selectively into the inner or the outer subunits of the intasome using a pair of IN variants—K120E and D273K—that co-assemble into hybrid

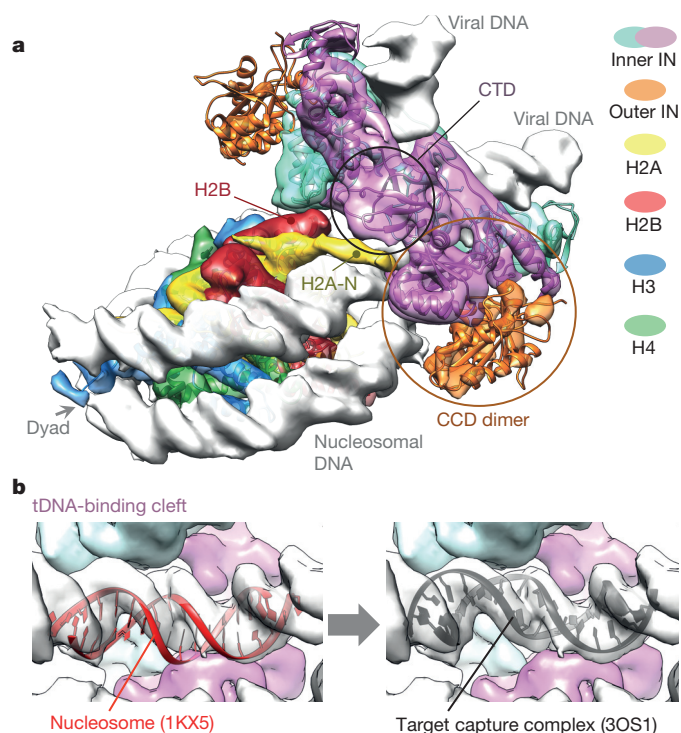


Figure 2 | Structure of the intasome–nucleosome complex. **a**, Segmented electron density map as semi-transparent surface with docked intasome and nucleosome structures shown as ribbons. H2B, the N-terminal tail of H2A (H2A-N), the CTD and one of the CCD dimers are indicated. **b**, Nucleosomal DNA within the tDNA-binding cleft of the intasome. DNA conformations as in available nucleosome structures (left) and as in the crystals of the PFV target capture complex (right) produce local electron density cross-correlation scores of 0.36 and 0.70, respectively. Protein Data Bank accessions are indicated in brackets.

intasomes through restricted localization to the inner and the outer subunit, respectively (Extended Data Fig. 6a–c). IN residues Gln 137, Lys 168 and Lys 159 are located in the vicinity of the contacts with the second gyre of nucleosomal DNA, while Pro 135, Pro 239 and Thr 240 approach the C-terminal helix of H2B (Fig. 3a). Substitutions at these positions affect the ability of the intasome to engage nucleosomes to various degrees. In particular, K168E in the inner and outer subunits, or the double substitution P135E/T240E in the inner subunits, grossly affected the interaction with recombinant and native human nucleosomes (Fig. 3b, c). The same substitutions reduced the ability of the intasome to integrate into the W601 nucleosome (Fig. 3d), while importantly preserving strand transfer activity into naked plasmid DNA (Extended Data Fig. 6d).

To assess the importance of the observed interactions under conditions of viral infection, we used a PFV vector system with a green fluorescent protein (GFP) reporter (Extended Data Fig. 7a)¹⁹. Initially, we mapped the genomic positions of 153,447 unique integration events of wild-type vector in human epithelial HT1080 cells. As a reference data set, we determined ~2.2 million integration sites using purified PFV intasomes and deproteinized human DNA. In agreement with published observations^{20,21}, PFV disfavoured transcription units: 32% of integration sites were found in RefSeq genes, which is ~16% below the gene-targeting frequency in the reference data set ($P = 10^{-300}$; Fig. 3e and Extended Data Table 1). Ranking transcription units by their activity in HT1080 cells, we discovered that PFV integration strongly contrasted with local transcriptional activity ($P = 10^{-300}$; Fig. 3e). Furthermore, integration sites accumulated within gene-poor and lamin-A/B-associated genomic regions²², indicating a strong bias towards condensed perinuclear heterochromatin (Fig. 3e and Extended Data Table 1).

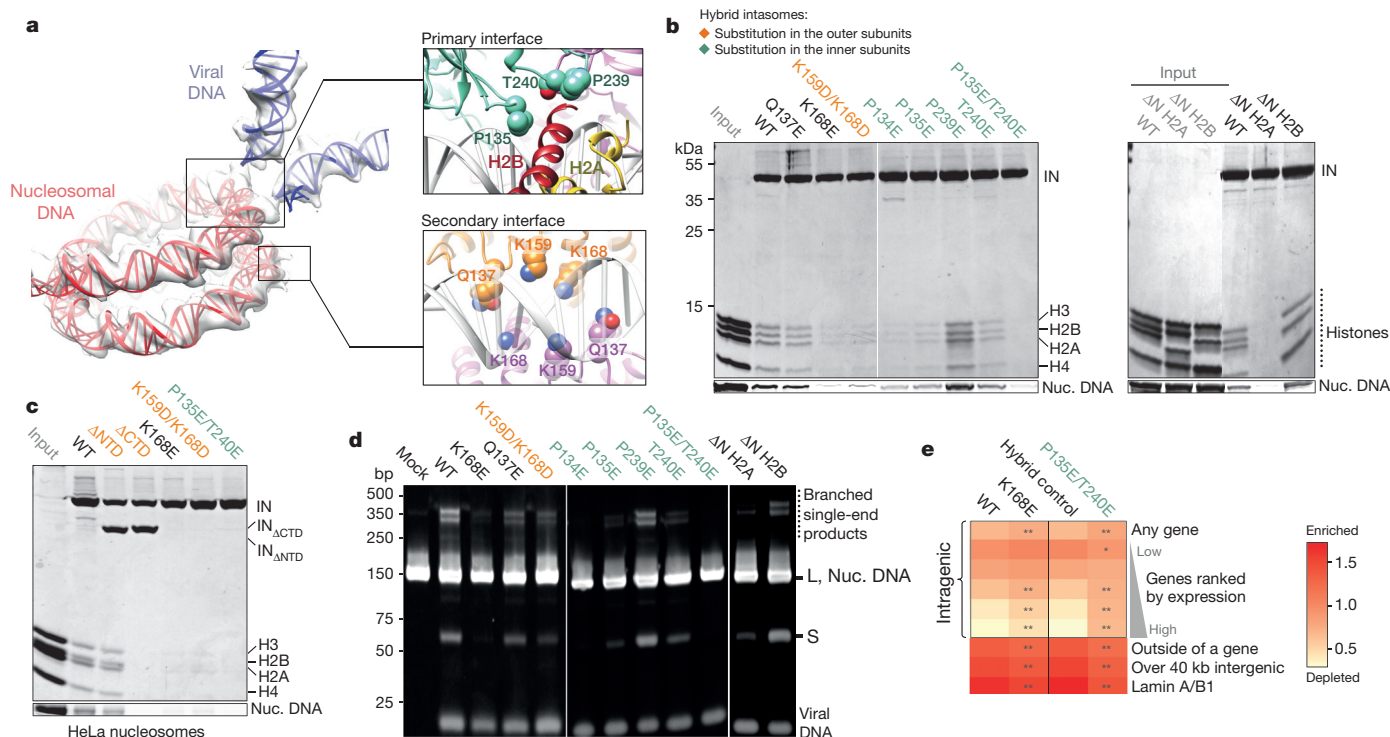


Figure 3 | Mutagenesis of the intasome–nucleosome interface. **a**, DNA components of the complex. The insets show pertinent details of the interface, with selected IN amino acid residues as spheres. **b**, Pull-down of H04 nucleosomes assembled with wild-type (WT) or variant (Δ N H2A or Δ N H2B, lacking residues 1–12 or 1–23, respectively) histones with biotinylated intasomes at 240 mM (left) or 290 mM NaCl (right). Substitutions in hybrid PFV intasomes were restricted to the outer subunits or inner IN subunits (indicated in orange and cyan font, respectively). Nuc. DNA, nucleosomal DNA. **c**, Pull-down of HeLa nucleosomes with biotinylated intasomes at 240 mM NaCl. **d**, Reactivity of intasome variants with W601 nucleosome assembled with wild-type or variant histones. DNA products (L and S),

separated by PAGE, were detected with GelRed. **e**, Heat map of relative integration frequencies. Each row represents a specific feature characteristic of the targeted genomic region. The colour scale indicates whether a feature is enriched (red) or depleted (light orange) compared with the *in vitro* control. Each column represents a PFV vector variant tested: wild type, K168E, hybrid control (vector produced using a combination of K120E and D273K/D185N/E221Q pcoP-POL packaging vectors encoding functional inner and catalytically incompetent outer IN subunits, respectively) or P135E/T240E (the substitutions introduced into the inner subunits in the hybrid control background). Asterisks indicate significant departures of the mutants from their respective controls (χ^2 test: * $P < 0.01$, ** $P < 10^{-5}$).

PFV vectors incorporating K168E IN or P135E/T240E IN as the inner subunit of the intasome displayed 2–4-fold defects in their ability to stably transduce HT1080 cells, while still strongly outperforming matched controls harbouring catalytically inert IN variants (Extended Data Fig. 7b–d). The residual infectivity allowed the mapping of 14,872 and 10,148 unique integration sites of K168E and P135E/T240E IN vectors, respectively. In accordance with their location outside of the tDNA-binding groove of the intasome, the mutations did not affect the weak sequence bias at the sites of integration (Extended Data Fig. 8). However, the mutations strongly corrupted the ability of the virus to discriminate against highly expressed genes, also leading to a modest but highly significant reduction in integration into gene-poor, lamina-associated regions (Fig. 3e and Extended Data Table 1). These observations are consistent with the importance of the intasome–nucleosome interactions observed in our structure for integration into condensed heterochromatin, normally preferred by PFV. Nucleosomes within transcriptionally active chromatin display the highest rate of remodelling and turnover, undergoing partial or complete disassembly to allow passage of RNA polymerase^{23,24}. Transient destabilization of the nucleosome arrays within highly expressed genes would provide a window of opportunity for the mutant viral intasomes.

In addition to the contacts involving structured regions of the intasome and the nucleosome, their flexible domains also contribute to the interaction. Segmentation of the cryo-EM map reveals a tubular density that probably represents the amino-terminal tail of H2A reaching out to the C-terminal domain (CTD) of one of the inner IN subunits (Fig. 2a). Concordantly, deletion of the H2A but not H2B N-terminal tail reduced intasome binding and strand transfer into nucleosomes

(Fig. 3b, d). Moreover, deletion of the N-terminal domains (NTDs) and CTDs, belonging to the outer subunits, which were unresolved in the previous crystal structures^{2,3,13,14,25} and in our averaged cryo-EM density (Fig. 2a), reduced the ability of the intasome to pull-down nucleosomes (Extended Data Fig. 2a). Further mutagenesis revealed a role of the outer CTDs (Fig. 3c), possibly through promiscuous interactions of the positively charged domains with nucleosomal DNA.

Using the nucleosomal structure as a landing platform may allow the viral machinery to sense epigenetic marks as well as provide a surface for recruitment of chromatin remodellers for disassembly of the post-catalytic strand transfer complex.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 1 February; accepted 15 April 2015.

Published online 10 June 2015.

- Li, M., Mizuuchi, M., Burke, T. R. Jr & Craigie, R. Retroviral DNA integration: reaction pathway and critical intermediates. *EMBO J.* **25**, 1295–1304 (2006).
- Hare, S., Gupta, S. S., Valkov, E., Engelmann, A. & Cherepanov, P. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464**, 232–236 (2010).
- Maertens, G. N., Hare, S. & Cherepanov, P. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468**, 326–329 (2010).
- Montaño, S. P., Pigli, Y. Z. & Rice, P. A. The mu transpososome structure sheds light on DDE recombinase evolution. *Nature* **491**, 413–417 (2012).
- Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
- Pryciak, P. M., Sil, A. & Varmus, H. E. Retroviral integration into minichromosomes *in vitro*. *EMBO J.* **11**, 291–303 (1992).

7. Pryciak, P. M. & Varmus, H. E. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* **69**, 769–780 (1992).
8. Pruss, D., Bushman, F. D. & Wolffe, A. P. Human immunodeficiency virus integrase directs integration to sites of severe DNA distortion within the nucleosome core. *Proc. Natl Acad. Sci. USA* **91**, 5913–5917 (1994).
9. Müller, H. P. & Varmus, H. E. DNA bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes. *EMBO J.* **13**, 4704–4714 (1994).
10. Wang, G. P., Ciuffi, A., Leipzig, J., Berry, C. C. & Bushman, F. D. HIV integration site selection: analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res.* **17**, 1186–1194 (2007).
11. Roth, S. L., Malani, N. & Bushman, F. D. Gammaretroviral integration into nucleosomal target DNA *in vivo*. *J. Virol.* **85**, 7393–7401 (2011).
12. Baller, J. A., Gao, J., Stamenova, R., Curcio, M. J. & Voytas, D. F. A nucleosomal surface defines an integration hotspot for the *Saccharomyces cerevisiae* Ty1 retrotransposon. *Genome Res.* **22**, 704–713 (2012).
13. Hare, S., Maertens, G. N. & Cherepanov, P. 3'-processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J.* **31**, 3020–3028 (2012).
14. Yin, Z., Lapkouski, M., Yang, W. & Craigie, R. Assembly of prototype foamy virus strand transfer complexes on product DNA bypassing catalysis of integration. *Protein Sci.* **21**, 1849–1857 (2012).
15. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
16. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
17. Dyda, F. *et al.* Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* **266**, 1981–1986 (1994).
18. Shaytan, A. K., Landsman, D. & Panchenko, A. R. Nucleosome adaptability conferred by sequence and structural variations in histone H2A–H2B dimers. *Curr. Opin. Struct. Biol.* **32**, 48–57 (2015).
19. Müllers, E. *et al.* Novel functions of prototype foamy virus Gag glycine-arginine-rich boxes in reverse transcription and particle morphogenesis. *J. Virol.* **85**, 1452–1463 (2011).
20. Trobridge, G. D. *et al.* Foamy virus vector integration sites in normal human cells. *Proc. Natl Acad. Sci. USA* **103**, 1498–1503 (2006).
21. Nowrouzi, A. *et al.* Genome-wide mapping of foamy virus vector integrations into a human cell line. *J. Gen. Virol.* **87**, 1339–1347 (2006).
22. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
23. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).
24. Schwabish, M. A. & Struhl, K. Asf1 mediates histone eviction and deposition during elongation by RNA polymerase II. *Mol. Cell* **22**, 415–422 (2006).
25. Gupta, K. *et al.* Solution conformations of prototype foamy virus integrase and its stable synaptic complex with U5 viral DNA. *Structure* **20**, 1918–1928 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the European Union FP7 HIVINNOV consortium grant 305137 (to P.C.), the US National Institute of General Medical Sciences P50 grant GM082251-06 (to P.C.) and the US National Institutes of Health R01 grant AI070042-08 (to A.N.E.). Data collection was in part funded by the Netherlands Centre for Electron Nanoscopy (NeCEN) by grants from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (project 175.010.2009.001) and by the European Union's Regional Development Fund through 'Kansen voor West' (project 21Z.014). We would like to thank L. Collinson, R. Carzaniga and Kirsty MacLellan-Gibson for EM access, R. Horton-Harpin for provision of HeLa cell pellets and assistance with tissue culture. We also thank F. Santoni, N. Sweeney and all our colleagues for helpful discussions.

Author Contributions D.P.M. analysed interactions of the PFV intasome and nucleosomes, discovered conditions to produce the stable intasome–nucleosome complex and prepared it for EM; L.R., D.P.M. and A.C. performed all EM work with the exception of cryo-EM grid preparation and screening, which was performed by L.R.; R.M. collected cryo-EM data; L.R. and A.C. determined the structure. S.H. designed the co-dependent K120E–D273K PFV IN pair; D.L. designed and provided wild-type PFV vector constructs; P.C. cloned PFV vector mutants; P.C. and P.L. carried out PFV infections; E.S. and A.N.E. developed the protocol and carried out sequencing of PFV integration sites; P.C. analysed integration site distributions.

Author Information The cryo-EM electron density map has been deposited in the Electron Microscopy Data Bank under accession number EMD-2992. Integration sites have been deposited in the NCBI Gene Expression Omnibus under accession number GSE67730. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.C. (alessandro.costa@crick.ac.uk) or P.C. (peter.cherepanov@crick.ac.uk).

METHODS

PFV intasome assembly. PFV IN was expressed in bacteria and purified as previously described²⁶. Synthetic DNA oligonucleotides used for intasome assembly were purified using high-performance liquid chromatography (HPLC) (Midland Certified). For pull-downs, activity assays and EM studies, PFV intasomes were assembled with stabilized processed U5 DNA (ref. 27) obtained by annealing oligonucleotides 5'-TGCGAAATTCATGACA (transferred strand) and 5'-ATTGTCATGGAATTTTCGCA. For experiments involving biotin pull-downs or in-gel fluorescent detection of strand transfer products, the transferred strand oligonucleotide was synthesized with 5' triethylene glycol biotin or fluorescein, respectively. To allow compatibility with the linker-mediated PCR protocols, a longer donor DNA construct spanning 47 bp of processed PFV U5 end and made by annealing oligonucleotides 5'-GATGTAATCCTTAGGATAATCA ATATACAAAATTCATGACA and 5'-ATTGTCATGGAATTTGTATAT TGATTATCCTAAGGAGTTACATC was used to generate *in vitro* genomic integration sites.

The intasomes were assembled according to published procedures^{2,27}, by dialysing 120 µM PFV IN and 50 µM donor DNA duplex combined in 500 mM NaCl and 50 mM BisTris propane-HCl, pH 7.45, against excess low-salt buffer containing 200 mM NaCl, 20 mM BisTris propane-HCl, pH 7.45, 2 mM dithiothreitol (DTT) and 25 µM ZnCl₂ for 16 h at 18 °C. For hybrid intasome assembly, 60 µM of each IN variant (K120E and D273K) was used. The intasomes were purified by size-exclusion chromatography through a 10/300 GL Superdex-200HR column (GE Healthcare) in 20 mM BisTris propane-HCl, pH 7.45, 320 mM NaCl and kept on ice for immediate use.

Preparation of native human mononucleosomes. Native human mononucleosomes were prepared according to established protocols²⁸ with minor modifications. HeLa cells were lysed in 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT, 0.5 mM phenylmethanesulfonylfluoride (PMSF), 10 mM HEPES-NaOH, pH 7.9, using a 40 ml Dounce homogenizer (Wheaton). Nuclei, harvested by gentle centrifugation, were digested with 0.5 U ml⁻¹ micrococcal nuclease (Sigma-Aldrich) in 4 volumes of 0.34 M sucrose, 3 mM CaCl₂, 60 mM KCl, 0.5 mM PMSF and 50 mM Tris-HCl, pH 7.5, for 10 min at 37 °C. The reaction was stopped with 5 mM EDTA, and the nucleosomes, extracted by addition of 0.5 M NaCl, were dialysed overnight at 4 °C against 650 mM NaCl, 2 mM EDTA, 1 mM β-mercaptoethanol and 20 mM HEPES, pH 7.4. Nucleosomes were isolated by size-exclusion chromatography through a Superdex-200 column operated in 650 mM NaCl, 1 mM EDTA and 20 mM Tris-HCl, pH 7.5. For the use in strand transfer assays (Fig. 1a), native nucleosomes, purified by size-exclusion chromatography and dialysed overnight against 100 mM NaCl, 1 mM EDTA and 20 mM Tris-HCl, pH 7.5, were separated by native 8% PAGE in a Model-491 preparative cell (Bio-Rad Laboratories). Nucleosomes were concentrated to 3 mg ml⁻¹ using Vivaspinn devices (GE Healthcare) and stored on ice.

Assembly of recombinant nucleosomes. DNA fragments for recombinant nucleosome assembly were generated by PCR using in-house produced Pfu polymerase and a DNA Engine Tetrad-2 thermal cycler (BioRad Laboratories). DNA products, pooled from four or eight 96-well PCR plates (100 µl reaction per well), were concentrated by precipitation with ethanol, re-dissolved in 1 mM EDTA and 5 mM Tris-HCl, pH 8.0, and injected into a 1 ml column packed with POROS-HQ 50 µm resin (Life Technologies). After extensive wash with salt-free buffer, DNA was eluted with a linear 0–2 M NaCl gradient in 5 mM Tris-HCl, pH 8.0. The fractions containing the PCR fragment were pooled and concentrated by ethanol precipitation. The procedure allows production of virtually any nucleosome-size DNA fragment on a mg scale. Human H2A, H3 and H4 histones, expressed in bacteria in inclusion bodies, were purified as previously described²⁹. Human H2B was expressed with an N-terminal hexahistidine tag, which was removed before histone octamer refolding. Nucleosomes, assembled by dialysis, were purified using native 8% PAGE and heat re-positioned by incubation at 37 °C for 30 min, as previously described (Extended Data Fig. 1a)²⁹.

Biotin pull-down assays. Ten micrograms biotinylated intasome was allowed to bind 10 µg recombinant or native nucleosomes in 700 µl of pull-down buffer (140–340 mM NaCl, 10% glycerol, 1 mM DTT, 0.1% Nonidet P-40 and 50 mM BisTris propane-HCl, pH 7.45), in the presence of 20 µl of streptavidin agarose (Life Technologies). After incubation for 2 h of end-over-rocking at 4 °C, the resin was washed in 5 changes of 700 µl pull-down buffer. Bound nucleoprotein complexes, dissociated by incubating the resin in 30 µl of 1.3× Laemmli SDS sample buffer at 37 °C for 5 min, were analysed by electrophoresis in 18% SDS–PAGE gels (Life Technologies). Proteins and DNA were detected by staining the gels with Coomassie instant blue and GelRed (Biotium), respectively.

To generate a library of nucleosomal DNA fragments enriched in tighter intasome binders, a biotin pull-down experiment setup with a tenfold excess of nucleosomes was carried out in the presence of 290 mM NaCl. The DNA component of recovered nucleosomal fraction was treated with S1 nuclease and calf intestinal

phosphatase (New England Biolabs). Blunt-ended and dephosphorylated DNA was incubated with Taq Polymerase to add deoxyadenosine to 3' ends for cloning into pCR4-TOPO (Life Technologies). Individual clones were sequenced using M13 forward primer.

Strand transfer assays. Reactions were carried out using intasomes purified by size-exclusion chromatography. A standard reaction contained 1.25 µg of intasome and 5 µg of nucleosome in 300 µl of 240 mM NaCl, 5 mM MgCl₂, 1 mM DTT, 4 µM ZnCl₂, 25 mM BisTris propane-HCl, pH 7.45. Strand transfer was allowed to proceed for 15 min at 37 °C, and the reaction was stopped by addition of 0.5% SDS and 25 mM EDTA. DNA products, deproteinized by digestion with 30 µg proteinase K at 37 °C for 1 h and ethanol precipitation, were separated in 4–12% TBE PAGE gels. Fluorescein-labelled DNA was detected using a Typhoon TRIO fluorescence scanner (GE Healthcare); non-labelled DNA was visualized by staining with GelRed. Strand transfer assays using naked supercoiled plasmid target DNA was done according to published procedures².

Nucleosome thermal denaturation assays. The heat denaturation assays were done in a CFX96 real-time detection system (Bio-Rad Laboratories) using Sypro Orange (Life Technologies, distributed as 5,000× stock solution). Five micrograms nucleosome in 150 mM NaCl, 25 mM Tris-HCl, pH 7.5, supplemented with 2.5× Sypro Orange in a final volume of 100 µl was heated from 20 to 95 °C with a 30-s hold every 0.5 °C, after which fluorescence was recorded (excitation at 523 nm and detection at 564 nm). Melting temperatures were determined using CFX manager software from the first derivative of the signal curve.

Preparation of intasome-nucleosome complex for EM. The complex, assembled by incubating 200 µg PFV intasome and 200 µg D02 nucleosome in 300 µl of 320 mM NaCl, 20 mM BisTris propane, pH 7.45, for 20 min at room temperature, was purified by size-exclusion chromatography over a Superdex-200HR 10/300 column in 320 mM NaCl, 25 mM BisTris propane-HCl, pH 7.45. Fractions containing the complex were immediately used to prepare EM grids.

Negative-stain single-particle analysis. Negative-stain EM grids were prepared as follows. Carbon was evaporated onto freshly cleaved mica using a Q150TE coater (Quorum Technologies) and baked for 2 h at 50 °C before floating onto 400-mesh copper grids (Agar Scientific). Dry continuous carbon grids were glow discharged for 30 s at 45 mA using a 100× glow discharger (Electron Microscopy Sciences). A 4 µl drop of sample was applied onto the glow-discharged grid immediately after elution from the gel filtration column. The grid was sequentially laid on top of four distinct 75 µl drops of a 2% (w/v) uranyl formate solution, stirred for 10 s, before blotting to dryness and being stored at room temperature before imaging. Negative stain grids were screened on a G2 Spirit LaB6 microscope (FEI) and data were collected on a JEOL-2100 LaB6 electron microscope (JEOL) operated at 200 kV. Images were recorded at a nominal magnification of ×50,000 on a UltraScan 4k × 4k CCD camera (Gatan), resulting in a 2.2 Å pixel size at the specimen level. A total of 135 micrographs were collected with a 1 to 2 µm defocus and using 20 e⁻ per Å². Contrast transfer function (CTF) estimation was performed with CTFFind3 (ref. 30) and micrographs were phase-flipped using Bsoft (ref. 31). Reference-free two-dimensional averages were calculated using routine MSA/MRA IMAGIC protocols³². The initial three-dimensional model was determined starting from a sphere, and further refined with multi-model projection-matching approaches using libraries from the EMAN2 and SPARX packages^{33,34} (Extended Data Fig. 9).

Cryo-EM structure determination. A 4 µl sample drop was applied to plasma-cleaned C-Flat grids (400 mesh CF-1/1, Electron Microscopy Sciences). After 1 min incubation, grids were double side blotted for 3.8 s in a CP3 cryo-plunger (Gatan) at 80% humidity and plunge frozen into –172 °C liquefied ethane. Cryo-grids were screened for ice quality using a 914 side-entry cryo-holder (Gatan) on a JEOL-2100 LaB6 operated at 200 kV and equipped with an UltraScan 4k × 4k CCD camera (Gatan). Cryo-grids were stored in liquid nitrogen and dry-shipped to the Netherlands Centre for Electron Nanoscopy (NeCEN). At NeCEN, cryo-grids were loaded into a Titan Krios electron microscope (FEI) for automated data collection with the EPU software (FEI) over a period of 3 days. Images were recorded at a nominal magnification of ×59,000 on a Falcon direct electron detector. One-thousand four-hundred and seventy-nine micrographs were recorded using a –1.5/–3.5 µm defocus range with an electron dose of 40 e⁻ per Å². A first cryo-EM map was calculated starting from a subset of 16,702 particles manually picked from 157 micrographs using Xmipp software³⁵. Multimodel refinement protocols described earlier were employed to solve a 19.7 Å resolution cryo-structure (based on 6,057 particles), using the 60 Å low-pass filtered negative stain volume as a starting model. A complete, 333,545-particle data set was generated after automated picking in Xmipp on the best 932 micrographs, selected after inspection of the power spectrum. Contrast transfer function was estimated using CTFFIND3. All further processing was performed within the RELION 1.2 environment³⁶. A first round of two-dimensional classification was performed to discard poorly averaging particles,

resulting in a cleaned 193,569-particle data set. A three-dimensional classification was subsequently performed using three classes and starting from the initial cryo-EM map; 83,500 particles belonging to the best three-dimensional class were selected and subjected to one further round of three-dimensional classification. This approach yielded an improved volume calculated from 53,887 particles. These particles were subsequently separated into 187 groups, on the basis of their refined intensity scale-factor, and used in a final three-dimensional refinement using the selected, 60 Å low-pass filtered three-dimensional class as a starting model. The density map was corrected for the modulation transfer function (MTF) of the Falcon detector and sharpened by applying a $-818 \text{ \AA}^{-2} B$ factor, estimated automatically with the RELION post-processing function (Extended Data Fig. 4g). Final resolution after post-processing was 7.84 Å, according to the 0.143 cut-off criterion (Extended Data Fig. 4d). The handedness of the map was verified by inspection of the duplex DNA density. UCSF Chimera was used for automated rigid-body docking and generating figures and videos³⁷.

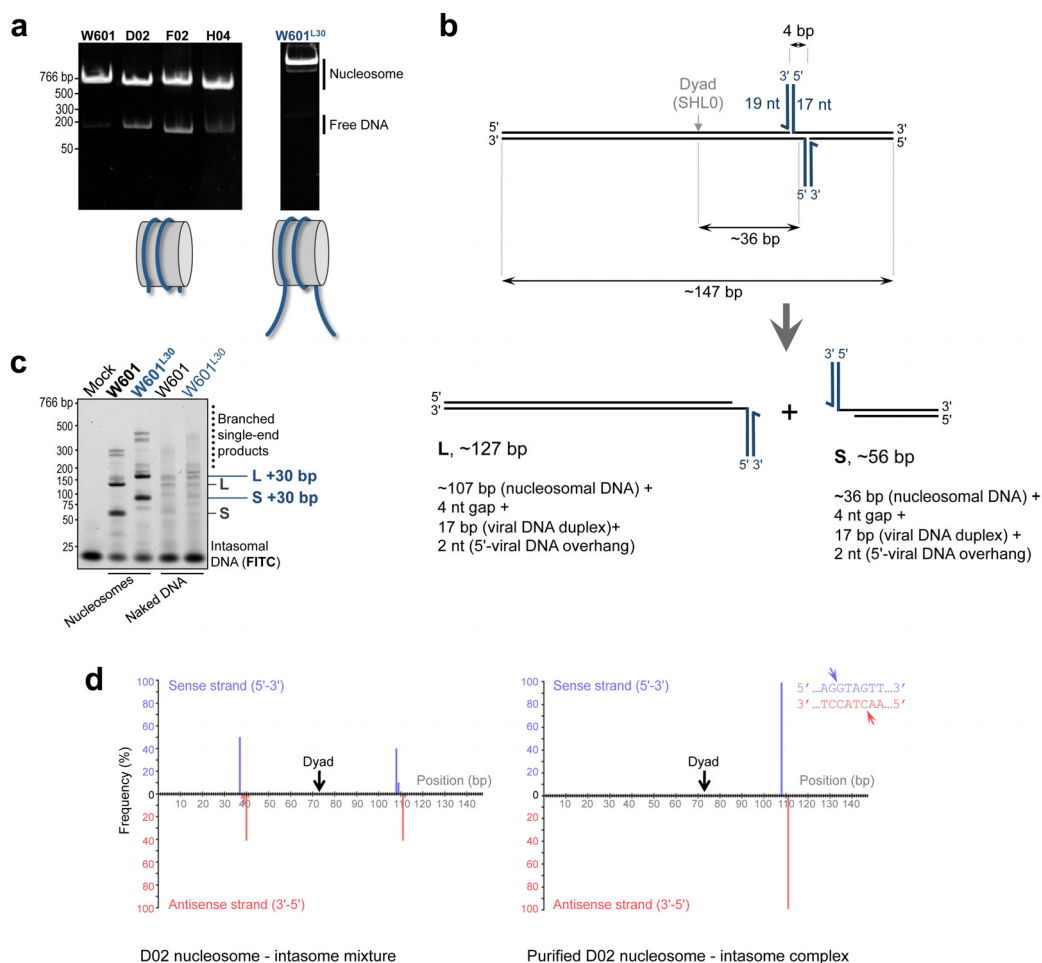
PFV vector infections and integration site analysis. Single-cycle viruses were produced by co-transfection of HEK293T cells (Cell Services, London Research Institute) with pMD9 (GFP reporter PFV vector) and codon-optimized foamy virus GAG, POL and ENV packaging constructs, as previously described^{19,38}. Hybrid PFV vectors and matched controls were pseudotyped with the less fusogenic simian foamy virus envelope³⁹, which allowed the use of higher virus inputs. Mutations were introduced into POL expression construct (pcoP-POL)¹⁹ by replacement of a BamHI/XbaI fragment spanning the IN coding region with respective mutant versions assembled by overlap-splicing PCR. Viruses were concentrated by centrifugation of cell-free supernatants through 20% sucrose cushion and re-suspended in PBS for immediate use. PFV GAG was detected in viral lysates using western blotting with rabbit polyclonal antibodies⁴⁰. HT1080 cells⁴¹, obtained from Cell Services of the London Research Institute, were cultured in DMEM supplemented with 10% fetal calf serum. The cells were infected with PFV vectors at a confluence of 20–30% and allowed to expand for 5–7 days to eliminate non-integrated viral DNA before cell sorting analyses and/or genomic DNA isolation. To generate *in vitro* integration sites, 10 µg deproteinized genomic DNA obtained from uninfected human cells was incubated with 0.8 µg PFV intasome in 1 ml of 125 mM NaCl, 10 mM MgCl₂, 2 mM DTT, 20 µM ZnCl₂, 25 mM BisTris propane, pH 7.45, for 30 min at 37 °C.

The protocols for linker-mediated PCR were adapted for amplification of PFV U5-genomic DNA junctions from published procedures^{42–44}. Briefly, genomic DNA isolated from infected HT1080 cells was digested with BfaI/BanII/CviQI and ligated to an asymmetric linker made by annealing synthetic oligonucleotides 5'-TAGTCCCTTAAGCGGAG (with 3' end blocked by an amino group modifier) and 5'-GTAATACGACTCACTATAGGGC NNNNNCTCCGCTTAAGGG AC (the string of random nucleotides comprise a serial number⁴³, not relevant to this study). PFV U5 integration sites were PCR-amplified using primers 5'-CAA GCAGAAG ACGGCATACGAGATCGGTCTCGGCATTCTGCTGAACCGC TCITCCGATC TGTAATACGACTCACTATAGGGC and 5'-GTGTGAAC TA CACTTATCTTAA ATGATG, followed by a nested PCR with primers 5'-CAA GCAGAAGACGGCATACGAGATCGGTCTCGGCATTCTGCTGA ACCGC TCITCCGATCTGTAATACGACTCACTATAGGGC and 5'-AATGATACG GCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTC CGATC TXXXXXCTTAAATGATGTAACCTCTTAGGATAATCAATATAC (the string of bold nucleotides marked with 'X' denotes the position of a 5-base index, which was unique for each integration site library analysed). The products were sequenced using MiSeq Illumina platform at the Dana-Farber Cancer Institute Molecular Biology Core Facilities.

The sequences of PFV U5-genomic DNA junctions were aligned to the hg19 version of the human genome (<http://genome.ucsc.edu/>) using BLAT⁴⁵ with settings -stepSize = 6, -minIdentity = 97 and -maxIntron = 0. The output was

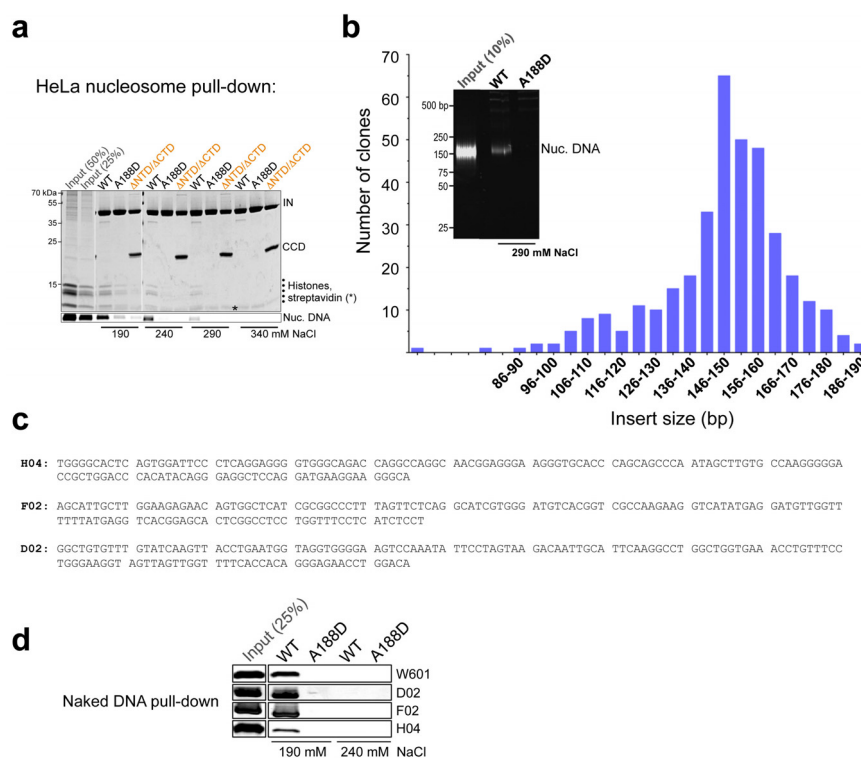
filtered using a custom Python script to retain only unique, high-quality matches starting with the first nucleotide downstream of the processed U5 sequence CAAATTCATGACA. The genomic coordinates were converted to the standard browser extensible data (BED) format, with each interval reporting the centre dinucleotide of a PFV integration site. Distributions of the integration sites with respect to genomic features (<http://genome.ucsc.edu/cgi-bin/hgTables>), HT1080-specific gene expression activity (Gene Expression Omnibus accession GSE58968)⁴⁶ and lamina-associated domains (GSE22428)⁴⁷, were analysed using BEDtools software suite⁴⁸. Nucleotide sequence logos were generated using WebLogo software (<http://weblogo.threeplusone.com/>)⁴⁹.

26. Valkov, E. *et al.* Functional and structural characterization of the integrase from the prototype foamy virus. *Nucleic Acids Res.* **37**, 243–255 (2009).
27. Hare, S. *et al.* Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *Proc. Natl Acad. Sci. USA* **107**, 20057–20062 (2010).
28. Schnitzler, G. R. Isolation of histones and nucleosome cores from mammalian cells. *Curr. Protoc. Mol. Biol.* **Chapter 21**, Unit 21 (2001).
29. Dyer, P. N. *et al.* Reconstitution of nucleosome core particles from recombinant histones and DNA. *Methods Enzymol.* **375**, 23–44 (2004).
30. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
31. Heymann, J. B. & Belnap, D. M. Bsoft: image processing and molecular modeling for electron microscopy. *J. Struct. Biol.* **157**, 3–18 (2007).
32. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).
33. Tang, G. *et al.* EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
34. Hohn, M. *et al.* SPARX, a new environment for Cryo-EM image processing. *J. Struct. Biol.* **157**, 47–55 (2007).
35. de la Rosa-Trevín, J. M. *et al.* Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* **184**, 321–328 (2013).
36. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
37. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
38. Heinkelstein, M. *et al.* Improved primate foamy virus vectors and packaging constructs. *J. Virol.* **76**, 3774–3783 (2002).
39. Stirnagel, K. *et al.* Differential pH-dependent cellular uptake pathways among foamy viruses elucidated using dual-colored fluorescent particles. *Retrovirology* **9**, 71 (2012).
40. Stange, A., Luftenegger, D., Reh, J., Weissenhorn, W. & Lindemann, D. Subviral particle release determinants of prototype foamy virus. *J. Virol.* **82**, 9858–9869 (2008).
41. Rasheed, S., Nelson-Rees, W. A., Toth, E. M., Arnstein, P. & Gardner, M. B. Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer* **33**, 1027–1033 (1974).
42. Matreyek, K. A. *et al.* Host and viral determinants for MxB restriction of HIV-1 infection. *Retrovirology* **11**, 90 (2014).
43. Chatterjee, A. G. *et al.* Serial number tagging reveals a prominent sequence preference of retrotransposon integration. *Nucleic Acids Res.* **42**, 8449–8460 (2014).
44. Schröder, A. R. *et al.* HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* **110**, 521–529 (2002).
45. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
46. Deyle, D. R. *et al.* A genome-wide map of adeno-associated virus-mediated human gene targeting. *Nature Struct. Mol. Biol.* **21**, 969–975 (2014).
47. Meuleman, W. *et al.* Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23**, 270–280 (2013).
48. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
49. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).



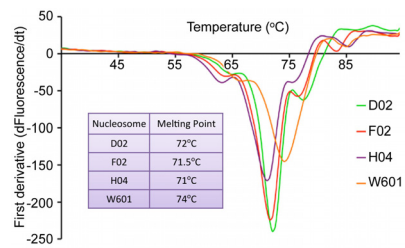
Extended Data Figure 1 | PFV integration into recombinant mono-nucleosomes. **a**, W601, D02, F02 and H04 nucleosome core particles (left) and W601 nucleosome with 30 bp tails mimicking linker DNA (W601^{L30}; right) were separated by native PAGE and detected by staining with ethidium bromide. **b**, Major products of PFV integration into a nucleosome core particle. Concerted integration of intasomal oligonucleotides (blue lines) into discontinuous tDNA (black lines) produces pairs of strand transfer products containing viral DNA mimics joined to tDNA fragments via 4 bp gaps. nt, nucleotides. **c**, PFV integration into nucleosome core particles (W601) and extended nucleosomes (W601^{L30}). Fluorescein-labelled intasomal DNA and

reaction products were separated by PAGE and detected by fluorescence scanning. Migration positions of the strand transfer products obtained with W601^{L30} nucleosome shift relative to those with the W601 core particle by ~30 bp. Thus, linker DNA does not appear to influence integration. **d**, Positions of integration events on D02 nucleosomal DNA before (left) or after (right) purification of the complex. The histograms show relative frequencies of integration events along the D02 DNA fragment into the top (blue bars) or bottom (pink bars) strands. The inset shows the nucleotide sequence at the preferred integration site; arrowheads indicate precise positions of the major integration events into the top and bottom strands of D02 DNA.

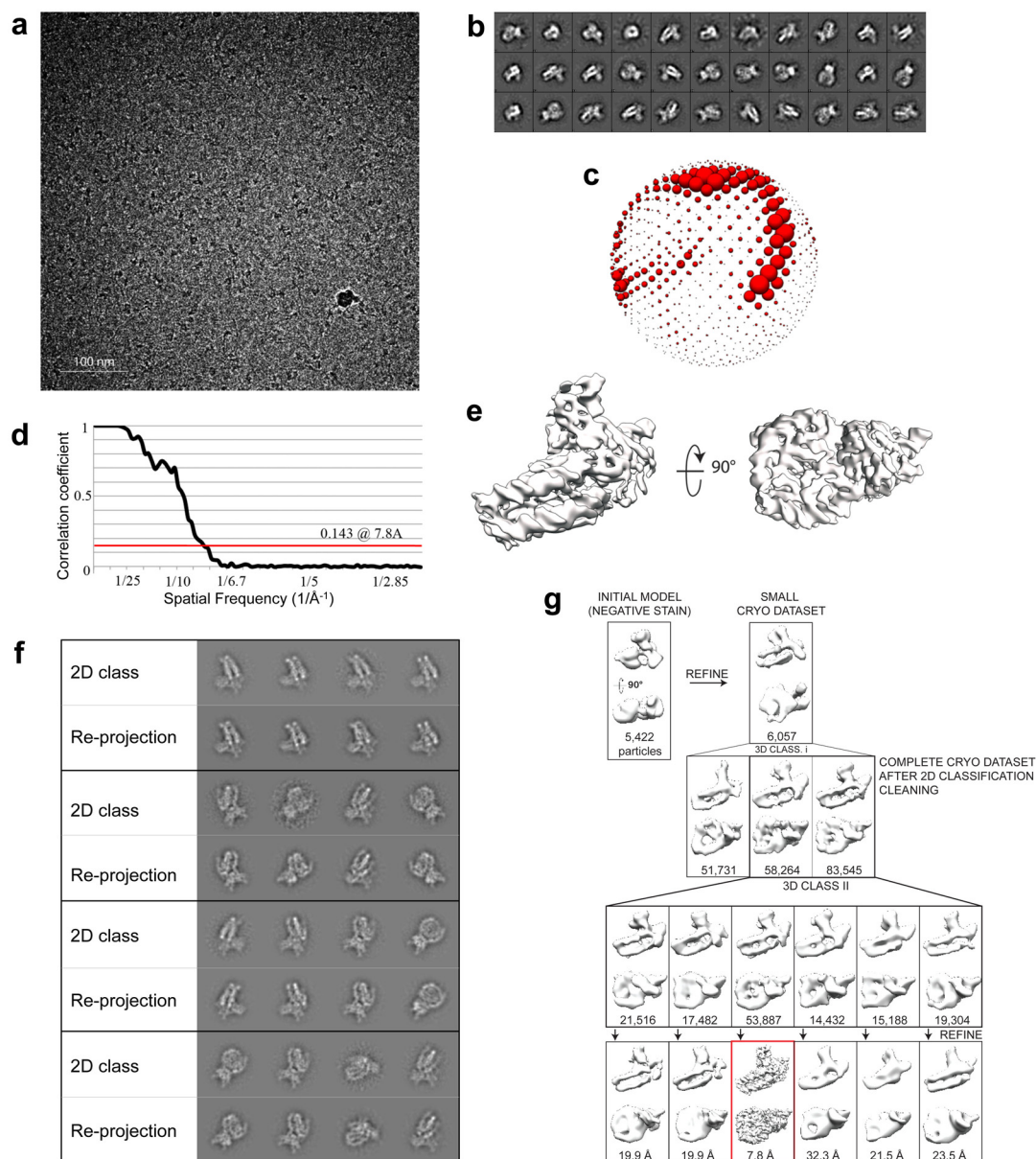


Extended Data Figure 2 | Pull-down of native nucleosomes and naked DNA by biotinylated PFV intasome. **a**, Mono-nucleosomes prepared by micrococcal nuclease digestion of HeLa cell chromatin were incubated with biotinylated intasomes under conditions of indicated ionic strength (190–340 mM NaCl). The intasomes used were wild type (WT), A188D or a hybrid intasome lacking the NTDs and CTDs on the outer subunits (indicated as Δ NTD/ Δ CTD; see main text and Extended Data Fig. 6a–c for details of hybrid intasome design). The intasome–nucleosome complexes were isolated on streptavidin agarose and separated by SDS–PAGE. Proteins and nucleosomal DNA were detected by staining with Coomassie blue and GelRed, respectively. Two leftmost lanes contained 50% and 25% of input nucleosomes, as indicated. Migration positions of protein sizes standards (kDa) are shown to the left of the gel. **b**, Isolation of HeLa nucleosomes preferentially binding to the PFV

intasome. Biotinylated wild-type or A188D intasomes were incubated with tenfold excess HeLa nucleosomes in the presence of 290 mM NaCl. Nucleosomal DNA recovered with wild-type intasome was cloned into a bacterial vector; the histogram depicts distribution of nucleosomal insert sizes obtained in this experiment. The inset shows separation of deproteinized nucleosomal DNA from 10% of input nucleosome material and from the fractions recovered with wild-type and A188D intasomes. **c**, Nucleotide sequences of three human DNA fragments (H04, F02 and D02) recovered with the intasome and used to assemble recombinant nucleosomes in this work. **d**, Naked W601 D02, F02 or H04 DNA was incubated with biotinylated wild-type or A188D intasomes in the presence of 190 or 240 mM NaCl, as indicated; DNA fractions recovered after pull-down on streptavidin beads were separated by PAGE and detected by staining with GelRed.

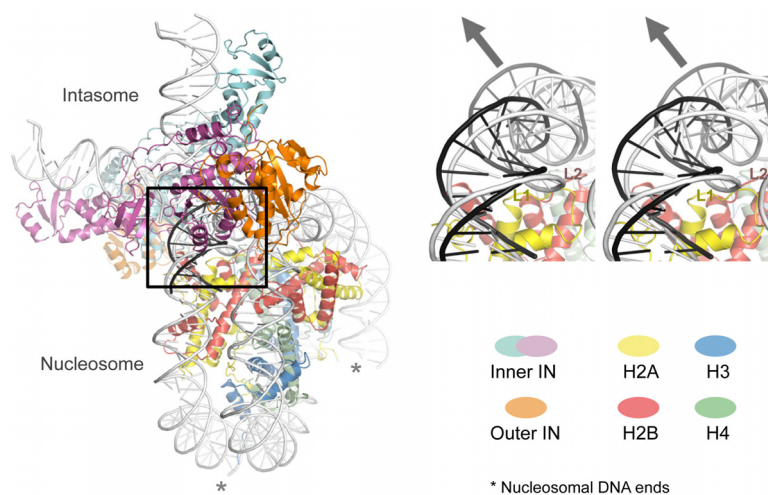


Extended Data Figure 3 | Thermal denaturation of recombinant nucleosomes. Derivative melt profiles of recombinant nucleosomes used in this study. The table in the inset shows experimentally determined melting temperatures.



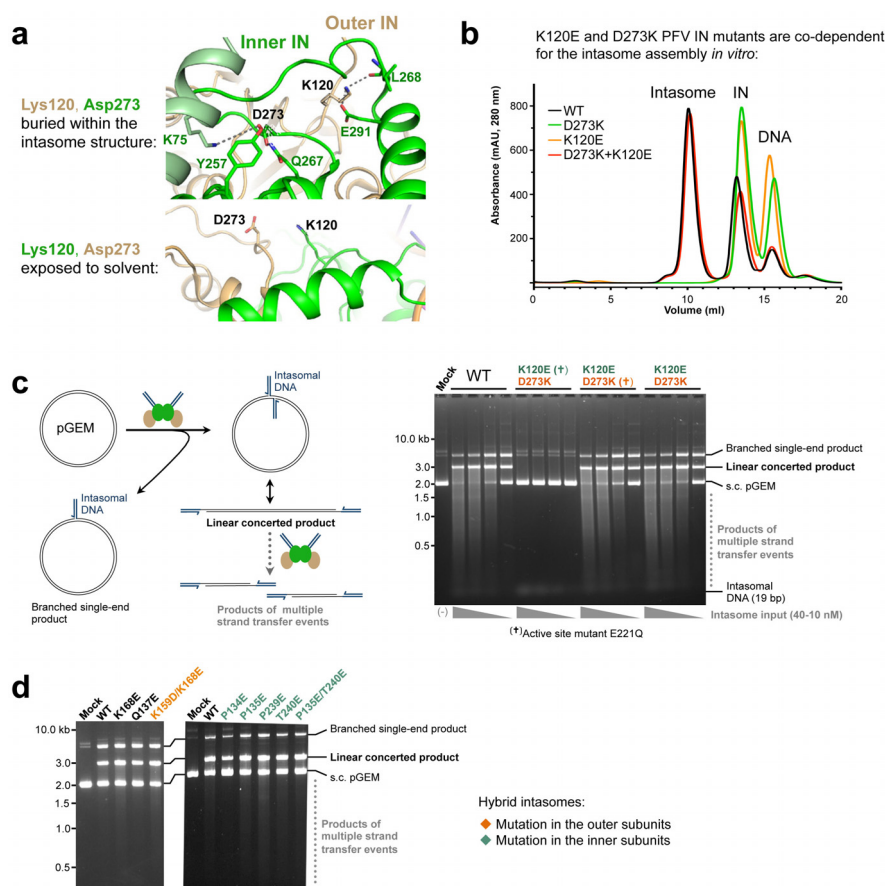
Extended Data Figure 4 | Overview of the cryo-EM data. **a**, Representative micrograph of frozen hydrated intasome-nucleosome complex. **b**, Two-dimensional class averages (phase-flipped only; box size 26 nm). **c**, Euler angle distribution of all particles included in the final three-dimensional reconstruction. Sphere size relates to particle number. **d**, Gold standard Fourier-shell correlation and resolution using the 0.143 criterion. **e**, Three-dimensional volume of the intasome-nucleosome complex refined with RELION. **f**, Match between reference-free two-dimensional class averages and three-dimensional re-projections of the cryo-EM structure. Two-dimensional class averages of fully contrast transfer function (CTF)-corrected particles are matched with the re-projections of the refined three-dimensional structure before map

sharpening (post-processing); 30–6 Å band-pass filter imposed. **g**, Overview of the three-dimensional classification and structure refinement. The initial negative stain structure was used for one round of structure refinement using a smaller cryo data set. The resulting map was used as a starting model for one round of three-dimensional classification (three classes) on a complete cryo data set. Particles from the two most populated three-dimensional classes were merged and used for one further round of three-dimensional classification (six classes). Each three-dimensional class was refined independently; the most populated three-dimensional class comprising 53,887 particles refined to 7.8 Å resolution.



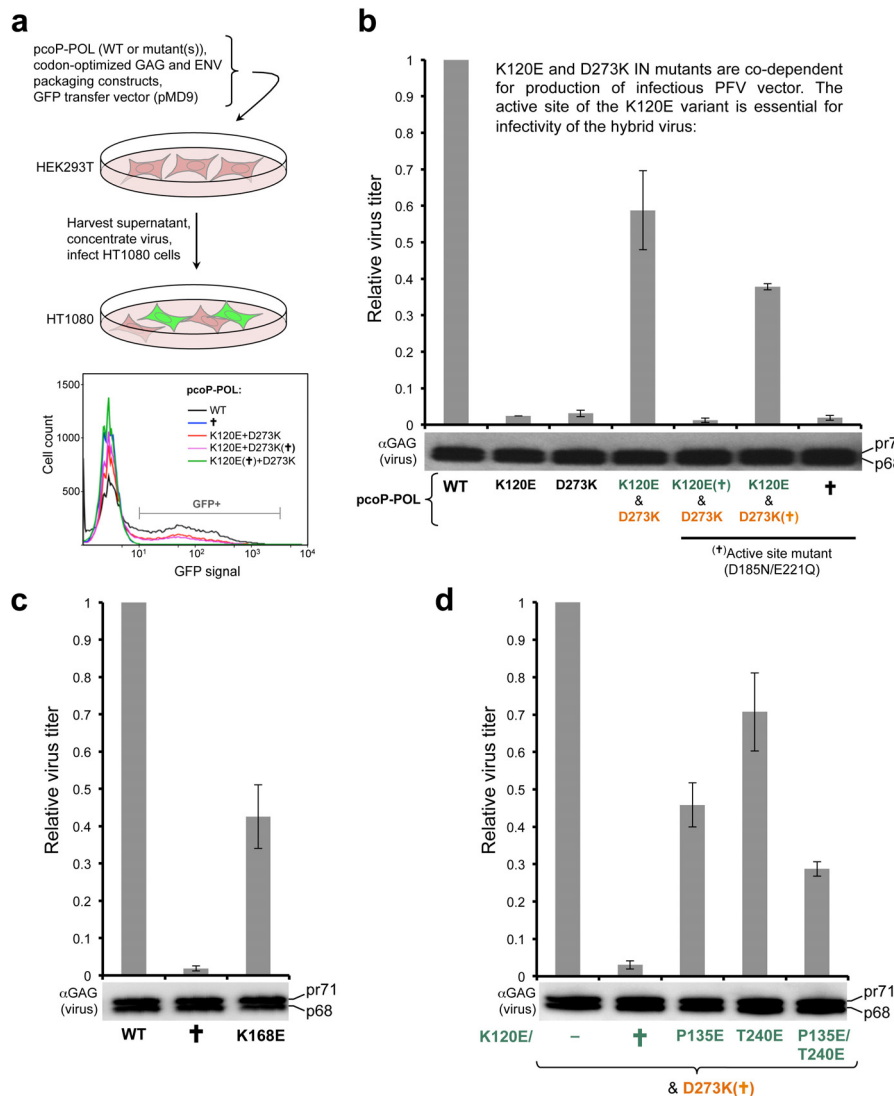
Extended Data Figure 5 | Nucleosomal DNA plasticity. Overview of the intasome–nucleosome complex structure (left) and a magnified stereo view of nucleosomal DNA engaged within tDNA binding cleft of the intasome (right). DNA conformations as in free nucleosomes (Protein Data Bank accession

1KX5) and as tDNA in complex with the PFV intasome (3OS1) are shown in light and dark grey, respectively; the arrowhead shows approximate direction of the DNA deformation. Asterisks indicate nucleosomal DNA ends.



Extended Data Figure 6 | Hybrid intasomes: structure-based design and validation *in vitro*. **a**, Views on the environment of Lys 120 and Asp 273 PFV IN residues within the intasome structure. Protein is shown as cartoons with side chains of selected amino acid residues shown as sticks; the cartoons and carbon atoms of the inner and outer IN chains are shown in green and light orange, respectively. Lys 120 of the outer and Asp 273 of the inner IN subunit are involved in a network of interactions; by contrast, Lys 120 of the inner and Asp 273 of the outer IN subunit are solvent-exposed. Consequently, IN mutants harbouring substitutions of Lys 120 or Asp 273 can only have a role in the inner or outer intasomal subunits, respectively; **b**, PFV IN mutants K120E and D273K are co-dependent for intasome assembly. Products of intasome assembly using wild-type (WT), D273K, K120E PFV IN or an equimolar mixture of D273K and K120E INs were separated by size-exclusion chromatography. Elution positions of the intasome, IN and free DNA are indicated. The assembly was successful with wild-type IN or with a mixture of the two IN mutants, but not with either of the IN variants separately. **c**, Validation of the hybrid intasome design. Left, possible types of strand transfer products obtained by reacting the intasome with circular DNA target

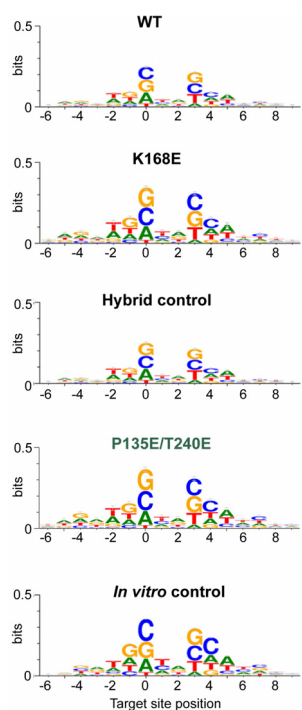
(pGEM, black lines). Full-site integration (strand transfer involving both intasomal DNAs, dark blue lines) results in a linear concerted product, which may be targeted by further strand transfer events. Half-site integration (strand transfer involving a single intasomal DNA end) results in a circular branched single-end product. Right, strand transfer assays using mutant intasomes and circular pGEM DNA target. The intasomes were assembled using wild-type IN or a mixture of K120E and D273K mutants, as indicated on top of the gel. IN variants marked with a cross additionally incorporated the E221Q amino acid substitution that disables the enzyme active site. Reaction products were separated by agarose gel electrophoresis. Intasomes were used at indicated concentrations; the leftmost lane contained a mock sample, which received no intasome. Migration positions of the reaction products, intasomal DNA and unreacted supercoiled (s.c.) pGEM are indicated to the right of the gel. As predicted, the strand transfer function of the hybrid intasome strictly requires the active site from the K120E (inner) IN subunit, but not the D273K (outer) subunit. **d**, Strand transfer activity of mutant intasomes on naked plasmid DNA. Mutations indicated in orange or green were restricted to the outer or inner subunits of the hybrid PFV intasome, respectively.



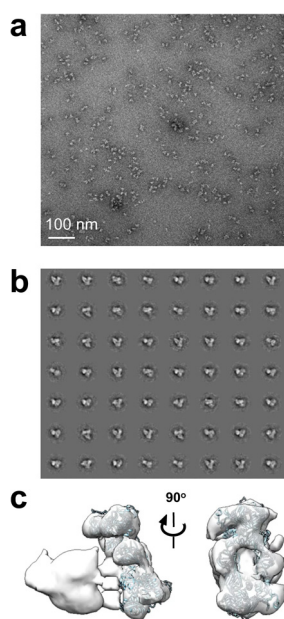
Extended Data Figure 7 | Infectivity of the mutant PFV vectors.

a, Schematic of the experiments. PFV vectors were produced in 293T cells transfected with DNA constructs encoding PFV GAG, POL and ENV, plus a GFP reporter transfer vector (pMD9). The virus, concentrated by centrifugation, was applied onto target HT1080 cells. Five days post-infection, the cells were analysed by FACS and/or used for isolation of genomic DNA and integration site sequencing. IN mutations were introduced into the packaging construct encoding POL (pcoP-POL). **b**, Validation of the hybrid intasome design in viral culture conditions. PFV GFP virus was produced using wild-type (WT), K120E, D273K POL packaging construct or a mixture of K120E and D273K mutants. The variants marked with a cross additionally contained a double point mutation inactivating the IN active site (D185N/E221Q). The graph and the western blot show mean relative infectivity and

GAG contents (pr71 and p68) of the resultant viruses, respectively. All infectivity experiments were done at least in triplicate, with two or more independent virus preparations; error bars represent standard deviations. The K120E and D273K IN mutants are co-dependent for production of infectious PFV vector, and the functional active site of the K120E IN component is essential for production of infectious hybrid virus. **c**, Relative infectivity of the PFV vectors harbouring wild-type, K168E, or active-site-dead D185N/E221Q (indicated with a cross) IN. **d**, Relative infectivities of hybrid viruses produced using D273K/D185N/E221Q (indicated as D273K with a cross) and K120E, K120E/D185N/E221Q (cross), K120E/P135E, K120E/T240E, and K120E/P135E/T240E. The western blots below the graphs show GAG (pr71 and p68) contents of the respective PFV vector preps.



Extended Data Figure 8 | Local nucleotide sequence biases at PFV integration sites. Nucleotide sequence preferences at PFV integration sites *in cellula* (wild type (WT), K168E, hybrid control and P135E/T240E) or *in vitro* displayed in the form of sequence logos. The heights of the logos correspond to the maximum information content at each position (maximum information content being 2 bits per base). Position 0 corresponds to the target nucleotide joined to the processed U5 PFV end.



Extended Data Figure 9 | Negative-stain EM analysis. **a**, Representative micrograph. **b**, Reference-free class averages. **c**, Three-dimensional electron density map of the intasome–nucleosome complex with a docked intasome structure. Note that DNA density is not recovered with negative-stain EM.

Extended Data Table 1 | Integration site preferences of PFV vector mutants

Dataset	Total sites	Intragenic*		± 5 kb TSS†		± 5 kb Lamin‡		Intergenic (≥40 kb)§		Gene density (Mbp ⁻¹)
		Number	%¶	Number	%	Number	%	Number	%	
<i>In vitro</i> [#]	2,212,496	1,050,686	47.5	213,822	9.7	880,461	39.8	713,115	32.2	10.3
WT	153,447	49,023	32.0	14,646	9.5	99,307	64.7	73,367	47.8	6.7
K168E	14,872	5,244	35.3 (<i>P</i> <10 ⁻¹⁴)	1,410	9.5 (<i>P</i> >0.5)	8,987	60.4 (<i>P</i> <10 ⁻²³)	6,572	44.2 (<i>P</i> <10 ⁻¹⁵)	7.4 (<i>P</i> <2.2·10 ⁻¹⁶)
Hybrid control**	41,243	12,862	31.2	3,447	8.4	28,251	68.5	20,320	49.3	6.3
P135E/T240E††	10,148	3,830	37.7 (<i>P</i> <10 ⁻³⁴)	801	7.9 (<i>P</i> >0.5)	5,701	56.2 (<i>P</i> <10 ⁻¹¹⁹)	4,361	43.0 (<i>P</i> <10 ⁻²⁸)	7.6 (<i>P</i> <2.2·10 ⁻¹⁶)

*Number and fraction of integration sites within RefSeq genes.

†Integration sites within 5 kb of a RefSeq transcription start site.

‡Integration sites within 5 kb of Lamin-A/B1 signal in HT1080 cells.

§Integration sites within extended intergenic regions (≥40 kb).

||Average number of genes within 500 kb of an integration site.

¶Percentage of sites within a feature; values for the mutant and paired control, which show statistically significant differences (*P* < 0.01) are shown in bold type. *P* values calculated for equivalence with wild-type or hybrid control using Chi-squared (for intragenic, ± 5 kb transcription start site, ± 5 kb lamin, and intergenic) or Mann–Whitney–Wilcoxon (for gene density) test are given in brackets. Bonferroni correction for multiple comparisons were applied to the *P* values.

#Reference data set containing integration sites obtained using purified PFV intasome and deproteinized human genomic DNA.

**Hybrid PFV vector harbouring K120E and D273K/D185N/E221Q IN mutants.

††Hybrid PFV vector harbouring K120E/P135E/T240E and D273K/D185N/E221Q IN mutants.

CORRIGENDUM

doi:10.1038/nature14511

Corrigendum: A new arboreal haramiyid shows the diversity of crown mammals in the Jurassic period

Xiaoting Zheng, Shundong Bi, Xiaoli Wang & Jin Meng

Nature **500**, 199–202 (2013); doi:10.1038/nature12353

In Fig. 2a of this Letter the tooth P3 should be horizontally flipped, as shown in Fig. 1 of this Corrigendum. We thank Z-X. Luo, T. Martin and C-F. Zhou for pointing out the error.



Figure 1 | Tooth P3, flipped horizontally.

ERRATUM

doi:10.1038/nature14567

Erratum: Mutant MHC class II epitopes drive therapeutic immune responses to cancer

Sebastian Kreiter, Mathias Vormehr, Niels van de Roemer, Mustafa Diken, Martin Löwer, Jan Diekmann, Sebastian Boegel, Barbara Schrörs, Fulvia Vascotto, John C. Castle, Arbel D. Tadmor, Stephen P. Schoenberger, Christoph Huber, Özlem Türeci & Ugur Sahin

Nature **520**, 692–696 (2015); doi:10.1038/nature14426

In this Letter, there were minor formatting errors in the poly-neo-epitope RNA in Fig. 3a, which occurred during the production process; this figure has been corrected in the online versions of the paper.

RETRACTION

doi:10.1038/nature14551

Retraction: HMGA2 functions as a competing endogenous RNA to promote lung cancer progression

Madhu S. Kumar, Elena Armenteros-Monterroso, Philip East, Probir Chakravorty, Nik Matthews, Monte M. Winslow & Julian Downward

Nature **505**, 212–217 (2014); doi:10.1038/nature12785

In this Letter, we reported that Hmga2 promotes lung cancer progression in mouse and human cells by operating as a competing endogenous RNA for the let-7 microRNA family. It has been brought to our attention that the cell lines used in the RNA sequencing (RNA-seq) experiment presented in Extended Data Fig. 4 of the Letter cannot be those specified in the figure legend. Moreover, these data cannot generate the Sylamer plots presented in this figure. The cell line substitution also casts doubt on which cells were used for experiments presented elsewhere in the paper. Although replication of other experiments in the Letter have not uncovered any further inconsistencies, given this uncertainty and the clear issues with Extended Data Fig. 4, we think it prudent to retract the paper pending more thorough investigation. We apologize for any adverse consequences that this may have caused.

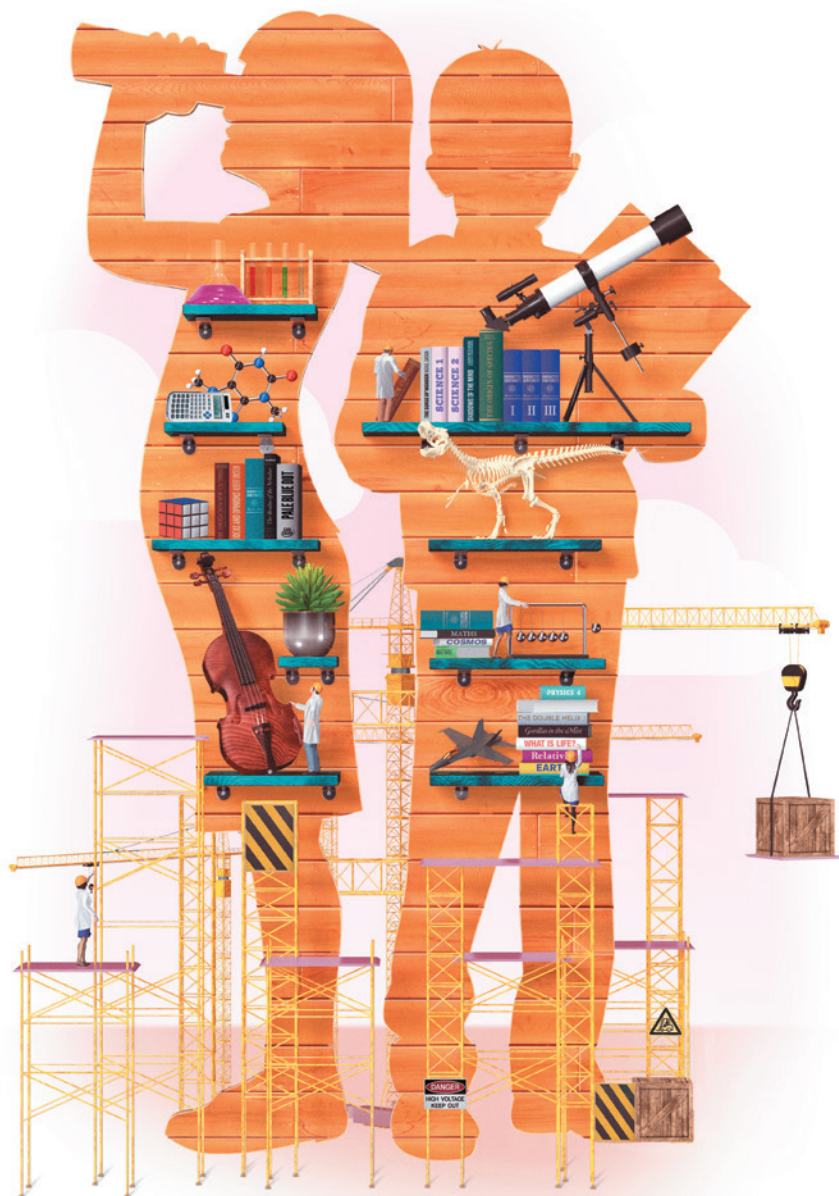
CAREERS

AFTER ACADEMIA Career awareness and preparedness go.nature.com/ube9gr

OUTREACH Five tips for talking to the media go.nature.com/wzspen

NATUREJOBS For the latest career listings and advice www.naturejobs.com

VASAVA



STEM EDUCATION

To build a scientist

Thought leaders across the globe answer one question: what is the biggest missing piece in how we educate scientists? Responses ranged from the practical to the philosophical.

PAUL NURSE

Expand across specialities

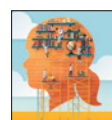
Director of the Francis Crick Institute, London

PhD programmes often lead to an increasing narrowness and specialization, which results in graduate students who are not sufficiently exposed to wider aspects of their subject and of related subjects. Looking outside the immediate interests of a thesis project can lead to real creative advances.

One way to expand thinking is to ensure that students have access to a series of inspirational speakers who will cover a wide range of scientific topics, with at least some who are more removed from their PhD focus. At the Francis Crick Institute, we will cover a wide range of biomedicine with truly inspirational speakers, but also look at other areas of science, such as high-energy physics, dark matter and aspects of biology, such as evolution and ecology, that are more distant from biomedicine.

Another suggestion is for what I call 'master classes', after the model of players of musical instruments. In science master classes, a group of graduate students would be exposed to a true expert, an excellent practitioner who would talk about doing science. I don't mean discussing the details of experiments, but discussing the broader questions: how do you do a satisfactory experiment, how do you do rigorous work, what is the nature of knowledge and so on.

The final suggestion is to broaden expectations. When students are three-quarters of the way through their graduate degree, they should be intensively mentored and urged to discuss their future careers. If they want to consider other careers, we need to build in a period of time — a few weeks — which they can use for short internships. We need to be honest, and acknowledge that not all of our students and postdocs will have a long-term career in basic research, but their education is still meaningful because they attain skill sets that they can take elsewhere — to enterprises that will profit from having scientists. We need to establish ►



THE 21ST CENTURY SCIENTIST
A *Nature* and *Scientific American*
special issue nature.com/stem

► a culture among advisers and investigators in which students who leave the academic pipeline are not considered 'failures'. They are making sensible choices and are to be cherished because they are taking science to other areas that will benefit from having them.

ATSUSHI SUNAMI

Broaden expertise across institutions

Professor at the National Graduate Institute for Policy Studies, Tokyo

As a nation, Japan needs more expertise in emerging fields such as brain science, cell engineering, data science and cybersecurity, but universities are still stuck in conventional scientific disciplines. We are asking universities to create programmes to represent these growing fields. Educational institutions need to cooperate to form a network of such programmes as they face the decline of Japan's university-age population and severe limitations on their resources.

Another urgent problem is how to encourage young scientists and engineers to go out and work with the best in the field and to gain the global connections that have become an essential aspect of science. Under changes to Japan's university system that have taken place over the past decade, many new positions are supported by competitive outside funding. This means that young scientists are hired on a fixed-term contract, which creates an insecure employment situation. Every 3–5 years, they look for another 3–5-year job. If we ask them why they do not go abroad to gain international experience, they say that they cannot risk losing the opportunity to secure another project in Japan. To resolve this, we are working to create international connections within our universities that will allow researchers to move to another country and back home again.

We also have to force change and diversity in the career track. In Japan's private sector, it is still rare for companies to hire PhD students and postdocs after they complete their training. In the past, it was almost customary to hire people directly from their undergraduate institution and route workers through their own training programmes, bypassing graduate education in exchange for lifetime employment. Universities can help to change the system: they can provide training and experience working in industry to mentor their PhD students and postdocs. To help to make this happen, we are introducing a scheme of cross-appointments of faculty-level experts in universities and companies. It will give trainees valuable skills and encourage companies to hire more PhD graduates and postdocs from universities.

JESSICA POLKA

Define purpose; demand decisions

Postdoctoral research fellow at Harvard Medical School, Boston, Massachusetts

What is missing from graduate education is a clear definition of its purpose. If graduate students are considered to be trainees, it behooves the funding agencies and everyone involved to make sure that their training is valuable to both society and the students. Graduate school is currently a research experience that is intellectually stimulating but not a clear stepping stone towards any career path. I question whether the graduate student–postdoc sequence is really necessary for training or whether it is a method of accruing credentials — and for getting science done at low cost. We should consider what benefit students gain from years four, five and even six versus their first three years. There needs to be a way to balance the needs of graduate students as students and not just as a research workforce.

To decide whether they will benefit from graduate school, people need to know where it may lead, and they need to stop thinking about faculty jobs as the probable end of the pipeline. The careers that people go into are diverse — many feel that they make use of their research training, but others do not. Mandates to create individual development plans for graduate students and to track their career outcomes would help to reveal what the job market actually looks like.

There should be more opportunities for

people to make conscious career decisions. For example, I think master's degrees should be more prevalent. People who take a master's after passing a qualifying exam should be viewed as making a reasonable decision about whether to pursue a PhD, and not for failing to continue as expected.

MICHAEL TEITELBAUM

Track PhDs after their degrees

Senior Research Associate at the Labor and Worklife Program, Harvard Law School, Cambridge, Massachusetts

For decades, aspiring young scientists in PhD programmes have been unable to get a good picture of what their career opportunities might be — not even of what recent graduates have experienced. That is a recipe for them to become disappointed, disheartened and potentially forced out of science. It is the responsibility of doctoral programmes to do their best to improve this situation.

Most graduate schools seem not to try very hard to keep track of their former PhD students and postdocs. They might know where their PhD graduates go for a postdoc, but not what they are doing 5–10 years on. Faculty members may know what their lab alumni are doing, but these data typically are not centrally assembled. That information, if universities compiled it as systematically as they do for those who earn their undergraduate degrees,



would be useful to prospective PhDs and postdocs who are thinking about their careers.

Universities should also consider limiting the length of the postdoc term. Many institutions have embraced formal limits — most commonly of five years — but these constraints can sometimes be sidestepped by a change in job title without a real change in role or prospects. Neither time constraints nor new job titles fixes the underlying problem of a lack of job options: the labour market for PhD scientists in most fields has not been robust. Understandably, they may want to continue for a sixth year in hopes that something will turn up, or stay for a seventh year and hope that they get that paper published in a top-tier journal.

If a postdoc wants to stay, if the principal investigator (PI) welcomes this and if there is research-grant money available, some ask why an arbitrary time limit should get in the way. But the dynamic is not working long term. Trainees need to understand that there could be diminishing career returns to opting for an extra year or two as a postdoc. Before they get to that point, PIs should be advising their PhD students and postdocs to broaden their skills beyond those typically taught in a PhD programme. Given the difficult current and prospective labour markets, well-advised PhD students and postdocs will probably realize that they need non-science professional and managerial skills if they wish to find attractive long-term careers that build on their scientific talents.

ROBERT TJIAN

Teach people management

President of the Howard Hughes Medical Institute, Chevy Chase, Maryland

My students and postdocs spend all their time focused on experiments, which is, of course, the top priority for young scientists who are building their careers. But something that we in the scientific community have not confronted very well is how to get them to focus on interacting productively with other people. Learning to manage teams and to work with others is going to become more important as science becomes more collaborative.

We are getting a little better at teaching students to write grant applications, but that is just a small part of running your own laboratory. The biggest part of leading a lab is getting the best work out of technicians, trainees and even colleagues. Typical graduate and postdoc programmes include little or no training in people management. I had to learn it by watching how my mentors ran their labs; there was no formal

management training of any sort. It took a while before I learned how to guide students without tearing down their self-confidence or how to motivate students in different ways depending on their personalities.

Outside master's programmes in business administration (MBAs), there is little training in leadership, how to form the right team and how to run it effectively. But how teams work together can really influence the way you do science. Regardless of whether things are going really well or everything is messed up, you, as the lab head, must keep cool and positive. You are the proverbial cheerleader, and getting depressed — and showing it — is rarely helpful.

Better training in lab and people management will also help lab heads to guide students to choose good problems and avoid getting overly enamoured with a specific model or system, and teach them to do experiments with rigour. Universities have to recognize that leadership training is a valuable lab skill, and they need to learn how to address it.

JARI KINARET

Practise the art of incisive questions

Director of the Graphene Flagship, Chalmers University of Technology, Gothenburg, Sweden

One of the issues that is not systematically covered in most graduate programmes is how to identify good research topics. Of course, there is no single way to do this — for one thing, it depends on what you regard as a good research topic, and opinions clearly differ. For every individual, the answer evolves as one acquires skills and experience, makes new contacts and so on, but some questions remain constant. Is this worth doing? Who cares if I or we succeed? Can I do it, either alone or with colleagues? What is the competition? Is this a one-off problem or is there a future in the area?

It is not clear whether the skill of choosing good topics can be taught, but it can clearly be learned: some researchers make the right choice more often than others, and it is hardly a talent that they have from birth. The first step is for supervisors and graduate students to discuss the choice — frequently, openly and critically. I think that this aspect of graduate studies is on the decline because many researchers are bound by their grants, which are usually written and decided before the student is hired, and many graduate students must execute a pre-defined plan within strict time constraints. Planning in advance is essential, of course, but training to set — and alter — topics for study is, or should be, an integral part of graduate studies.

JO HANDELSMAN

Match training to job trends

Associate director for science at the White House Office of Science and Technology Policy, Washington DC

Because academic jobs are scarce, some analysts have proposed reducing the number of trainee positions in science, technology, engineering and mathematics (STEM). But this argument errs in its assumption that STEM students are — and should be — trained exclusively for faculty positions at research universities.

It is true that only a small proportion of those who start STEM doctoral degrees from US institutions today will go on to attain faculty positions. In biology, for example, fewer than 8% of new PhD students do so. Although that statistic might look alarming, it does not reflect the growing employment needs and opportunities that exist outside of traditional academia.

Today, the United States actually needs more, not fewer, PhD graduates in STEM fields. We must abolish the idea that these people will aim solely for academic research posts. More than 98% of STEM PhD graduates are employed, and in diverse careers. Furthermore, faculty positions are no longer the top career goal of many graduate students. A 2011 survey at the University of California, San Francisco, for instance, found that its graduate students are increasingly eager to manage research labs, direct education programmes, write, make public policy, start companies and teach at small universities. Few of these keen students, however, receive training in the skills necessary for non-conventional careers.

Graduate education in STEM should evolve to meet these needs. Courses in pedagogy, science writing, entrepreneurship or administration offered either on campuses or by professional societies would equip PhD students to confront the broad scientific job market.

The incorporation of more diverse educational experiences into US graduate training need not lengthen the time commitment. At the University of Wisconsin–Madison, for instance, some STEM graduate students have been required to do a three-month internship in industry or government. The internship did not affect time to degree, perhaps because the experience strengthened students' focus and motivation.

If graduate training were redesigned to better prepare graduate students for non-academic research careers, would they pursue more-varied career opportunities, and more confidently? Would they be more satisfied with graduate school? It's worth finding out. ■

OUTPATIENT

A rude awakening.

BY DAN STOUT

It was definitely a migraine.

The agony clamped down on both temples, and the light from behind the curtain shot daggers through my eyelids. I twisted over to cover my head with a pillow and felt a sudden breeze up my backside.

I sat up, squinting, a hospital gown tugging at my throat. I had no idea what had happened to me. My last memory was of being in my lab, slipping on my sensor headdress and wiring it to the neural monitors.

Pushing the assistance buzzer, I rocked back and forth, trying to keep the migraine at bay. No nurse answered, and eventually I gave up. When I stood, I staggered, a stranger in my own body.

I stumbled out into the hall, relieved to see a familiar logo on the directional signs. I was still in St Anne's, the hub of my work, where Kim Stanley and I were pioneering Spatial Resonance Neurology — the expansion of the brain's network into the space around it, building awareness beyond our bodies.

The halls were jammed with patients looking just as confused as me. Apparently some were dealing with even worse headaches than I was, as they leaned against walls, gripping their temples or succumbing to the nausea and vomiting on the floor. The overwhelmed staff ran back and forth. No one paid me any attention.

I picked up a white technician's coat from a chair at the nurse's station. I'd had enough of my rear end being exposed. As I put it on, the collar flipped up. Even with a decade of practice I'd never quite figured out how to keep those things flat. I glanced around, one eye shut against the pain of my headache, and tried to figure out what was going on. So many people with signs of headache and nausea. *Gas leak?* There was no odour of

natural gas. *Carbon monoxide?* The hospital had CO detectors in every hall, but no alarms were

sounding. My cell phone would be in my office. I could call 911 and get outside.

Down one floor, having taken the stairs so I could bypass the yelling crowd at the elevator lobby, I reached my office. It had been

me, and I doubled over. Gasping, I made myself look back up at them.

Behind Kim and the woman, a teenage girl sat perched on a stool. She wore a hospital gown and squinted as if pained by the light. As I watched, she reached out and grabbed Kim's arm.

"It's me." The pleading note in her voice was heart-breaking. "It's Ellen."

There was a crash, and an obese man in a hospital gown stumbled through the doors. He showed clear signs of recent surgery.

"Kim," he said. "Something went wrong. I woke up in someone else's..." He trailed off as he stared at the woman next to Kim. "Oh, God," he said.

There was a spike of pain as my migraine raged back into full force. I raised a hand to massage my temple and saw the ID bracelet on my wrist, name and room number printed on treated plastic. My name was apparently Carol Jones.

Over my shoulder I could hear shuffling feet, a growing chorus of "Kim... please, Kim," as more and more

patients pressed into the lab. I did my best to ignore the occasional cry of "It's Ellen," as they made my stomach knot and the wave of nausea rise again.

To distract myself I tried to do some maths, remembering the range of our devices. I guessed at the population density of San Diego and tried to calculate just how many people would now flip up their collars and prefer their coffee with cream, just the way I liked it. I finally gave up, not really knowing if it mattered anymore. I covered my eyes, both from the harsh fluorescent glare of the lights and because I didn't want to look again at the too familiar woman standing next to Kim. Eyes shielded I rocked back and forth, trying futilely to hide from the migraine that I knew would only get worse. ■

Dan Stout loves to write about the things that terrify and inspire him. His fiction has appeared in Plan B Magazine and Mad Scientist Journal. Visit him at www.DanStout.com.



ILLUSTRATION BY JACEY